PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Commissioner
US Department of Commerce
United States Patent and Trademark
Office, PCT
2011 South Clark Place Room
CP2/5C24
Arlington, VA 22202
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 21 May 2001 (21.05.01)	
International application No. PCT/GB00/03576	**Applicant's or agent's file reference** JPD/P100087PCT
International filing date (day/month/year) 18 September 2000 (18.09.00)	**Priority date** (day/month/year) 17 September 1999 (17.09.99)
Applicant WOOLFSON, Derek, N. et al	

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International Preliminary Examining Authority on:

11 April 2001 (11.04.01)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO **34, chemin des Colombettes** **1211 Geneva 20, Switzerland**	Authorized officer Zakaria EL KHODARY
Facsimile No.: (41-22) 740.14.35	Telephone No.: (41-22) 338.83.38

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	**FOR FURTHER ACTION** see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.	
JPD/P100087PCT		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/GB 00/ 03576	18/09/2000	17/09/1999
Applicant		
UNIVERSITY OF SUSSEX		

This International Search Report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This International Search Report consists of a total of 3 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language,** the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

[ ] the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing :

[ ] contained in the international application in written form.

[ ] filed together with the international application in computer readable form.

[X] furnished subsequently to this Authority in written form.

[X] furnished subsequently to this Authority in computer readble form.

[X] the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

[X] the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished

2. [ ] **Certain claims were found unsearchable** (See Box I).

3. [ ] **Unity of invention is lacking** (see Box II).

4. With regard to the **title,**

[ ] the text is approved as submitted by the applicant.

[X] the text has been established by this Authority to read as follows:

PROTEIN STRUCTURES AND PROTEIN FIBRES

5. With regard to the **abstract,**

[X] the text is approved as submitted by the applicant.

[ ] the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. 1a

[ ] as suggested by the applicant.

[X] because the applicant failed to suggest a figure.

[ ] because this figure better characterizes the invention.

[ ] None of the figures.

Claims

1. A protein structure comprising a plurality of first peptide monomer units arranged in a first strand and a plurality of second peptide monomer units arranged in a second strand wherein the first and second monomer units comprise the heptad repeat motif (abcdefg) and/or the hendecad repeat motif (abcdefghijk), and wherein a pair of asparagines, arginines, lysines or other complementary residues in the "a" position on at least one pair of corresponding first and second monomer units ensures that the first strand and the second strand form a staggered parallel heterodimer coiled coil structure.

2. A protein structure according to claim 1, wherein a first peptide monomer unit in the first strand extends beyond a corresponding second peptide monomer unit in the second strand in the direction of the strands.

3. A protein structure according to any one of claims 1 to 2 in which at least one charged amino acid residue of a first peptide monomer unit is arranged to attract an oppositely-charged amino acid residue of a second peptide monomer unit.

4. A protein structure according to claim 3 in which the charged amino acid residue is in an end portion of the first peptide monomer unit which extends beyond the corresponding second peptide monomer unit in the second strand.

5. A protein structure according to any one of the preceding claims in which at least one strand consists solely of first or second peptide monomer units respectively.

6. A protein structure according to any one of the preceding claims wherein one or more of the other "a" positions of the first and second monomer units is a hydrophobic residue.

7. A protein structure according to claim 6, wherein the hydrophobic residue is selected from isoleucine or valine.

8. A protein structure according to any one of the preceding claims having a leucine at one or more of the "d" positions of the first and second monomer units.

9. A protein structure according to any one of the preceding claims having oppositely-charged or otherwise complementary residues at positions g and e of respective monomer units.

10. A protein structure according to claim 9 in which the oppositely-charged residues are glutamic acid and lysine residues or arginine and aspartic acid residues, or synthetic derivatives of these amino acid residues.

11. A protein structure according to any preceding claim in which the structure is stabilised by pairs of asparagine, arginine, lysine or other complementary residues provided by corresponding first and second peptide monomer units.

12. A protein structure according to any preceding claim which is arranged to form a tubular structure.

13. A protein structure according to claim 12 in which the repeat motifs are offset by two or more amino acid positions in sequence whereby the peptide monomer units form a cylinder.

14. A protein structure according to any preceding claim in which the first and second peptide monomer units have the sequence.

a) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1) and

b) KIRALKAKNAHLKQEIAALEQEIAALEQ (SAF-p2) respectively; or

c) KIAALKQKIAALKQEIDALEYENDALEQ (SAF-p1A) and

d) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C) respectively; or

e) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1C) and

f) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C) respectively.

15. A peptide monomer unit for use in preparing a protein structure the peptide monomer unit having an amino acid sequence selected from:

a) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1);

b) KIRALKAKNAHLKQEIAALEQEIAALEQ (SAF-p2);

c) KIAALKQKIAALKQEIDALEYENDALEQ (SAF-p1A);

d) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C) and

e) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1C).

16. A protein structure according to any one of claims 1 to 14 or a peptide monomer unit according to claim 15 wherein at least one amino acid residue is derivatised.

17. A branching self-assembling fibre comprising two or more protein structures according to any one of claims 1 to 11, coupled together to form a T-shaped conjugated structure.

18. The branching self-assembling fibre of claim 17, wherein at least one of the protein structures comprises one or more central cysteine residues, and at least one other protein structure comprises a terminal cysteine residue.

19. A method of producing protein structures, the method comprising providing a mixture of first and second monomer units which associate to form a protein structure according to any one of claims 1 to 14, wherein the first and second monomer units comprise the heptad repeat motif (abcdefg) and/or the hendecad repeat motif (abcdefghijk).

20. A method according to claim 19 in which the protein structure is derivatised.

21. A method according to claim 19 or 20 in which the protein structure is stabilised by cross-linking.

22. A protein fibre produced by an association of protein structures according to any one of claims 1 to 14.

23. A kit for making a protein structure, the kit comprising first and second peptide monomer units which associate to form a protein structure according to any one of claims 1 to 14 or a protein fibre according to claim 22, wherein the first and second monomer units comprise the heptad repeat motif (abcdefg) and/or the hendecad repeat motif (abcdefghijk).

24. A two dimensional grid comprising a protein structure according to any one of claims 1 to 14 or a protein fibre according to claim 22.

25. A three dimensional matrix comprising a protein structure according to any one of claims 1 to 14 or a protein fibre according to claim 22.

26. A matrix according to claim 25 which is arranged to assemble in solution.

27. A matrix according to claim 25 or claim 26, wherein one or more binders is fused to the protein structure, wherein the one or more binders are aligned to give high avidities for one or more target entities.

28. A matrix according to any one of claims 25 to 27 which is arranged to bind one or more target entities.

29. A matrix according to claim 28 which is arranged to bind viruses.

30. A method of forming a matrix according to any one of claims 25 to 29 in which a mixture of separate first and second monomer units is provided, wherein the first and second monomer units comprise the heptad repeat motif (abcdefg) and/or the hendecad repeat motif (abcdefghijk) and are caused to associate to form a plurality of protein structures according to any one of claims 1 to 14, wherein the protein structures assemble to form a three-dimensional matrix.

31. A method according to claim 30 in which the matrix is formed *in situ*.

32. A method for controlling the production of a synthetic polymers comprising assembling a protein structure in accordance to any one of claims 1 to 14 in association with the polymer.

33. A method according to claim 32 in which the protein structure is removed after synthesis of the polymer.

34. A tip for use in Atomic Force Microscopy comprising a protein structure according to any one of claims 1 to 14.

REPLACED ART 34 AMD

...nciple of using "sticky ends" is well developed in molecular biology for assembling DNA (S. J. Palmer *et al* (1998) *Nucleic Acids Res.* **26**, 2560), and has been used to design intricate DNA crystals (E. Winfree *et al* (1998) *Nature* **394**, 539). However, to our knowledge, our application of sticky end-directed molecular assembly to peptides is new; although we do note that head-to-tail packing of helices has been observed in recently solved crystal structures for two designer peptides (N. L. Ogihara *et al* (1997) *Protein Sci.* **6**, 80; G. G. Prive *et al* (1999) *Protein Sci.* **8**, 1400). These were helical peptides that crystallised with their helical ends in contact so as to form pseudo-continuous helices in the solid state. In other words they formed "blunt-ended" arrangements.

According to one aspect of the invention there is provided a protein structure comprising a plurality of first peptide monomer units arranged in a first strand and a plurality of second peptide monomer units arranged in a second strand, the strands preferably forming a coiled-coil structure, and in which a first peptide monomer unit in the first strand extends beyond a corresponding second peptide monomer unit in the second strand in the direction of the strands. The protein structures of the invention have numerous advantages. For example, relatively long protein fibres can be formed with little material - 1 µl of a 100 µM solution of the peptide monomers may provide enough material to form 10 m of fibre 50 nm thick.

At least one charged amino acid residue of the first peptide monomer unit may be arranged to attract an oppositely-charged amino acid residue of the second peptide monomer unit. Preferably, the charged amino acid residue is in an end portion of the first peptide monomer unit, which extends beyond the corresponding second peptide monomer unit in the second strand. At least one strand may consist solely of first or second peptide monomer units respectively i.e homogenous strands. Heterologous strands are also contemplated. The peptide monomer units may comprise a repeating structural unit. Preferably, the repeating structural unit comprises a heptad repeat motif, having the pattern:

```
hpphppp
abcdefg
```

Preferably, the repeat may include isoleucine or asparagine at position a and leucine at position d. Other repeats (e.g hendecads - abcdefghijk) and amino acid compositions may also be used (see WO99/11774).

Fig. 7 is an electromicrograph showing fibres which have been derivatised through the inclusion of flurophores; and

Fig. 8 shows amino acid sequences designed to form blunt-ended heterodimers.

1) Peptide Design and Synthesis

Various peptide monomer units were designed as described above. The monomers and capping peptides (designed to complement the sticky ends of the monomers so as to produce flush, or blunt ends and, so, arrest longitudinal fibre assembly) are set out in Table 1:

Claims

1. A protein structure comprising a plurality of first peptide monomer units arranged in a first strand and a plurality of second peptide monomer units arranged in a second strand in which a first peptide monomer unit in the first strand extends beyond a corresponding second peptide monomer unit in the second strand in the direction of the strands.

2. A protein structure according to claim 1 in which the strands together form a coiled coil structure.

3. A protein structure according to claim 1 or 2 in which at least one charged amino acid residue of a first peptide monomer unit is arranged to attract an oppositely-charged amino acid residue of a second peptide monomer unit.

4. A protein structure according to claim 3 in which the charged amino acid residue is in an end portion of the first peptide monomer unit which extends beyond the corresponding second peptide monomer unit in the second strand.

5. A polypeptide structure according to any preceding claim in which at least one strand consists solely of first or second peptide monomer units respectively.

6. A protein structure according to any preceding claim in which the peptide monomer units comprise a repeating structural unit.

7. A protein structure according to claim 6 in which the repeating structural unit comprises a heptad repeat motif (abcdefg).

8. A protein structure according to claim 6 in which the repeating structural unit comprises a hendecad repeat motif (abcdefghijk)

9. A protein structure according to claim 6 having isoleucine or asparagine at position a and leucine at position d.

10. A protein structure according to claim 6 having valine or leucine at positions a and d respectively.

11. A protein structure according to any one of claims 7 to 10 having oppositely-charged or otherwise complementary residues at positions g and e of respective monomer units.

12. A protein structure according to claim 11 in which the oppositely-charged residues are glutamic acid and lysine residues or asparagine and aspartic acid residues, or synthetic derivatives of these amino acid residues.

13. A protein structure according to any preceding claim in which the structure is stabilised by pairs of asparagine, arginine, lysine or other complementary residues provided by corresponding first and second peptide monomer units.

14. A protein structure according to any preceding claim which is arranged to form a tubular structure.

15. A protein structure according to claim 14 in which the peptide monomer units are offset by two or more amino acid positions in sequence whereby the peptide monomer units form a cylinder.

16. A protein structure according to any preceding claim in which the first and second peptide monomer units have the sequence:

a) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1) and

b) KIRALKAKNAHLKQEIAALEQEIAALEQ (SAF-p2) respectively; or

c) KIAALKQKIAALKQEIDALEYENDALEQ (SAF-p1A) and

d) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C) respectively; or

e) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1C) and

f) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C) respectively.

17. A peptide monomer unit for use in preparing a protein structure the peptide monomer unit having an amino acid sequence selected from:

a) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1);

b) KIRALKAKNAHLKQEIAALEQEIAALEQ (SAF-p2);

c) KIAALKQKIAALKQEIDALEYENDALEQ (SAF-p1A);

d) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C);

e) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1C) ; and

d) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C).

18. A protein structure or peptide monomer unit according to any preceding claim in which at least one amino acid residue is derivatised.

19. A method of producing protein structures, the method comprising providing a mixture of first and second peptide monomer units which associate to form a protein structure according to any preceding claim.

20. A method according to claim 19 in which the protein structure is derivatised.

21. A method according to claim 19 or 29 in which the protein structure is stabilised by cross-linking.

22. Protein fibres produced by an association of protein structures according to any one of claims 1 to 3 or a method according to claim 19, 20 or 21.

23. A kit for making protein structures, the kit comprising first and second peptide monomer units which associate to form a protein structure according to any one of claims 1 to 13 or protein fibres according to claim 22.

24. A two dimensional matrix comprising a protein structure according to any one of claims 1 to 13 or protein fibres according to claim 22.

25. A three dimensional grid comprising a protein structure according to any one of claims 1 to 13 or protein fibres according to claim 21.

26. A matrix according to claim 25 which is arranged to assemble in solution.

27. A matrix according to claim 25 or 26 which is arranged to bind a target entity.

28. A matrix according to claim 27 which is arranged to bind viruses.

29. A method of forming a matrix according to any one of claims 25 to 28 in which a mixture of separate first and second monomer units is provided and are then caused to associate to form a protein structure in accordance with the invention, an accumluation of such protein structures assembling in turn to form a three dimensional matrix.

30. A method according to claim 29 in which the matrix is formed *in situ*.

31. A method for controlling the production of a synthetic polymers comprising assembling a protein structure in accordance to any one of claims 1 to 16 in association with the polymer.

32. A method according to claim 31 in which the protein structure is removed after synthesis of the polymer.

33. A tip for use in Atomic Force Microscopy comprising a protein structure according to any one of claims 1 to 16.

INTERNATIONAL SEARCH REPORT

Interna[redacted]pplication No
PCT/[redacted]0/03576

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/00 C07K1/113 G11B9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, WPI Data, PAJ, CHEM ABS Data, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 712 366 A (KAPLAN DAVID L ET AL) 27 January 1998 (1998-01-27) column 6, line 35-57; figures 6A,6B ---	1-15, 18-30
X	WO 96 11947 A (GOLDBERG EDWARD B) 25 April 1996 (1996-04-25) page 7-9; figures 3,4 page 25-26 --- -/--	1-6,14, 15, 18-20, 22-30

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

°A° document defining the general state of the art which is not considered to be of particular relevance

°E° earlier document but published on or after the international filing date

°L° document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

°O° document referring to an oral disclosure, use, exhibition or other means

°P° document published prior to the international filing date but later than the priority date claimed

°T° later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

°X° document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

°Y° document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

°&° document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 January 2001	17/01/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Cervigni, S

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Interna[illegible]l Application No
PCT/[illegible]00/03576

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 5712366	A	27-01-1998	NONE		
WO 9611947	A	25-04-1996	US	5877279 A	02-03-1999
			AU	689662 B	02-04-1998
			AU	3829695 A	06-05-1996
			BR	9509487 A	30-09-1997
			CA	2202474 A	25-04-1996
			CN	1168676 A	24-12-1997
			EP	0785946 A	30-07-1997
			HU	77683 A	28-07-1998
			JP	10508194 T	18-08-1998
			NZ	295046 A	25-03-1998
			US	5864013 A	26-01-1999

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 00/03576

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/00 C07K7/113 G11B9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, WPI Data, PAJ, CHEM ABS Data, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 712 366 A (KAPLAN DAVID L ET AL) 27 January 1998 (1998-01-27) column 6, line 35-57; figures 6A,6B ---	1-15, 18-30
X	WO 96 11947 A (GOLDBERG EDWARD B) 25 April 1996 (1996-04-25) page 7-9; figures 3,4 page 25-26 --- -/--	1-6,14, 15, 18-20, 22-30

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 January 2001	17/01/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Cervigni, S

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	KOJIMA SHUICHI ET AL: "FIBRIL FORMATION BY AN AMPHIPATHIC ALPHA-HELIX-FORMING POLYPEPTIDE PRODUCED BY GENE ENGINEERING." PROCEEDINGS OF THE JAPAN ACADEMY SERIES B PHYSICAL AND BIOLOGICAL, vol. 73, no. 1, 1997, pages 7-11, XP000971780 1997 ISSN: 0386-2208 abstract	1-15,19, 22-30
X	W A PETKA ET AL: "REVERSIBLE HYDROGELS FROM SELF-ASSEMBLING ARTIFICIAL PROTEINS" SCIENCE,AAAS. LANCASTER, PA,US, vol. 281, 17 June 1998 (1998-06-17), pages 389-392, XP002149253 abstract; figure 2	1-13, 18-20, 23-30
A	KOHN W D ET AL: "De novo design of alpha-helical coiled coils and bundles: models for the development of protein-design principles" TRENDS IN BIOTECHNOLOGY,NL,ELSEVIER, AMSTERDAM, vol. 16, no. 9, September 1998 (1998-09), pages 379-389, XP004173181 ISSN: 0167-7799	

2

Information on patent family members

International Application No
PCT/GB 00/03576

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 5712366	A	27-01-1998	NONE		
WO 9611947	A	25-04-1996	US	5877279 A	02-03-1999
			AU	689662 B	02-04-1998
			AU	3829695 A	06-05-1996
			BR	9509487 A	30-09-1997
			CA	2202474 A	25-04-1996
			CN	1168676 A	24-12-1997
			EP	0785946 A	30-07-1997
			HU	77683 A	28-07-1998
			JP	10508194 T	18-08-1998
			NZ	295046 A	25-03-1998
			US	5864013 A	26-01-1999


OIPE
JC76
MAR 18 2002
PATENT & TRADEMARK OFFICE

REC'D 28 DEC 2001
WIPO PCT

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

Applicant's or agent's file reference	FOR FURTHER ACTION	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
JPD/P100087PCT		
International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/GB00/03576	18/09/2000	17/09/1999
International Patent Classification (IPC) or national classification and IPC C07K14/00		
Applicant UNIVERSITY OF SUSSEX et al.		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 8 sheets, including this cover sheet.

   ☒ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of 8 sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☒ Priority
   - III ☐ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☒ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☐ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☒ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
11/04/2001	21.12.2001
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Moonen, P Telephone No. +49 89 2399 8538

Form PCT/IPEA/409 (cover sheet) (January 1994)

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/GB00/03576

I. Basis of the report

1. With regard to the elements of the international application *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17))*:
Description, pages:

1,2,4-7,9-24	as originally filed			
8	as received on	10/01/2001	with letter of	08/01/2001
3,3a	with telefax of	05/12/2001		

Claims, No.:

1-34	with telefax of	05/12/2001

Drawings, sheets:

1/10-10/10	as originally filed

Sequence listing part of the description, pages:

1-5, filed with the letter of 09.11.2000

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item

These elements were available or furnished to this Authority in the following language: , which is:

- ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).
- ☐ the language of publication of the international application (under Rule 48.3(b)).
- ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

- ☐ contained in the international application in written form.
- ☐ filed together with the international application in computer readable form.
- ☐ furnished subsequently to this Authority in written form.
- ☐ furnished subsequently to this Authority in computer readable form.
- ☒ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.
- ☒ The statement that the information recorded in computer readable form is identical to the written sequence

listing has been furnished.

4. The amendments have resulted in the cancellation of:

   ☐ the description, pages:
   ☐ the claims, Nos.:
   ☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

   (Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

II. Priority

1. ☐ This report has been established as if no priority had been claimed due to the failure to furnish within the prescribed time limit the requested:

   ☐ copy of the earlier application whose priority has been claimed.

   ☐ translation of the earlier application whose priority has been claimed.

2. ☒ This report has been established as if no priority had been claimed due to the fact that the priority claim has been found invalid.

   Thus for the purposes of this report, the international filing date indicated above is considered to be the relevant date.

3. Additional observations, if necessary:
   see separate sheet

IV. Lack of unity of invention

1. In response to the invitation to restrict or pay additional fees the applicant has:

   ☐ restricted the claims.

   ☐ paid additional fees.

   ☐ paid additional fees under protest.

   ☐ neither restricted nor paid additional fees.

2. ☐ This Authority found that the requirement of unity of invention is not complied and chose, according to Rule 68.1, not to invite the applicant to restrict or pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rules 13.1, 13.2 and 13.3 is

☐ complied with.

☒ not complied with for the following reasons:
see separate sheet

4. Consequently, the following parts of the international application were the subject of international preliminary examination in establishing this report:

☒ all parts.

☐ the parts relating to claims Nos. .

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-34
	No:	Claims	
Inventive step (IS)	Yes:	Claims	1-11, 14-15 and 17-25
	No:	Claims	12-13, 16, and 26-34
Industrial applicability (IA)	Yes:	Claims	1-34
	No:	Claims	

2. Citations and explanations
see separate sheet

VIII. Certain observations on the international application

The following observations on the clarity of the claims, description, and drawings or on the question whether the claims are fully supported by the description, are made:
see separate sheet

Reference is made to the following documents:

D1: US-A-5 712 366 (KAPLAN DAVID L ET AL) 27 January 1998 (1998-01-27)
D2: WO 96 11947 A (GOLDBERG EDWARD B) 25 April 1996 (1996-04-25)
D3: KOJIMA SHUICHI ET AL: 'FIBRIL FORMATION BY AN AMPHIPATHIC ALPHA-HELIX-FORMING POLYPEPTIDE PRODUCED BY GENE ENGINEERING.' PROCEEDINGS OF THE JAPAN ACADEMY SERIES B PHYSICAL AND BIOLOGICAL, vol. 73, no. 1, 1997, pages 7-11,1997
D4: W A PETKA ET AL: 'REVERSIBLE HYDROGELS FROM SELF-ASSEMBLING ARTIFICIAL PROTEINS' SCIENCE,AAAS. LANCASTER, PA,US, vol. 281, 17 June 1998 (1998-06-17), pages 389-392
D5: Biochemistry **39** (June/August 2000) 8728-34; intermediate document

The document **D5** was not cited in the international search report. A copy of the document has not been provided, as the document is known to the inventors.

Re Item II
Priority

1. The present application claims the priority date of the priority application GB9922013.9, designated P1, having as filing date 17.09.1999. The subject-matter of P1 is more limited in scope than the internationally filed application.

   It is for example noted that present claims 12-13, 16 and 26-34, and examples 6 and 8-9 are missing in the priority document. The subject-matter of said claims is therefore only entitled to the filing date of 18.09.2000, making e.g. D5 available for citing as a prior art document.

Re Item IV
Lack of unity of invention

2. Given the partial lack of right to priority and the disclosure **D5** (referring to the sticky-end assembly of a designed peptide fiber) it is considered that the subject-matter of e.g. claims 26-28 is not necessarily linked to that of claim 1.
   A single general inventive concept (referred to in Rule 13 PCT and the PCT

Preliminary Examination Guidelines Ch.III, 7) is therefore not recognisable in the absence of a common, special technical feature.

Re Item V

Reasoned statement under Rule 66.2(a)(ii) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

3. The present application concerns the assembly of designed peptides with sticky-ends, in contrast to supramolecular assemblies made up of units having blunt ends (Ogihara et al.).

D1 is a patent publication (also considered to be the closest prior art) disclosing assemblies of dimers having also sticky-ends (see Figure 5 concerning the interaction between the oppositely sequences of Fig. 4A and 4B, Figures 6A and B, column 6 concerning recognition elements at the N- or C-termini to align all the dimers in a "head-to-tail" orientation within a growing fibril, and the claims concerning assemblies of coiled coils with charged heptad subunits).

New claim 1 contains presently the additional features of originally filed claims 6-8 and claim 13; it specifes that in the "a" position at least one pair of corresponding first and second monomer units a pair of Asn, Arg or Lys *or other complementary residues* is present to ensure that the first strand and the second strand form a staggered parallel heterodimer coiled coil structure. It is however noted that the peptide monomers specified in claim 15 all refer to Asn in one "a" position and that the the wording *"or other complementary residues"* is not clearly defined. Novelty is therefore recognised (when at least *or other complementary residues* does not refer to hydrophobic residues) for claim 1 over D1.

4. The other relevant available documents are:
D2 discloses in particular self-assemblies polypeptides to produce protein nanometer structures; this patent application does not relate to coiled coils made up of units with heptad repeats (D2 concerns ß-sheets), but it does concern the tinkering with peptide sequences for better binding in the assemblies, e.g. see page 8 "Structural units" referring to rods (strands) having positively and negatively charged groups or protrusions built in for specific binding to other units

and figures 2-3 (in particular the closed brickwork sheets having overlapping surfaces).
D3 is disclosing fibril formation (four helix bundle) by α-helix-forming polypeptide having Ala at a particular position in the heptad repeat motif; no reference is made to "sticky-ends".
D4 concerns reversible hydrogels from self-assembling recombinant proteins containing terminal leucine zipper domains (comprising heptad repeats and forming coiled coils agggregates; protein **3** contains two helix units of 42 residues).

5. **Inventive step:** None of these documents is prejudicial to the novelty of claim 1 or any other claim and novelty is therefore acknowledged (Article 33(2) PCT). The involvement of an inventive step (Article 33(3) PCT) for claims 1-11, 14-15 and 17-25 entitled to the priority date of 17.09.1999 is also ackowleged, as no document alone or any combination of available documents suggested the forming of the protein structures as claimed.

An inventive step is however not recognised for the claims not entitled to the priority date: D5 is in this case the closest prior art.
The present application does not satisfy the criterion set forth in Article 33(3) PCT because the subject-matter of **claim 16** does not involve an inventive step (Rule 65(1)(2) PCT): the feature of derivatization is known from the prior art (see e.g. WO 99/11774, also mentioned in the present description).
D5 refers to the possibility of higher-order assemblies (page 8733, right column) as well as D1 (see the bottom of column 6). The combination of D5 with the teaching of e.g. D5 makes therefor the subject-matter of **claims 12-13 and 26-31** obvious to the skilled person.
The subject-matter of **claims 32-34** appears to be based on general knowledge of the skilled person, and is therefore also obvious to the skilled person.

Re Item VIII
Certain observations on the international application

6. Claim 1 now refers to a protein structure with at least one pair of corresponding first and second monomer units ensuring that the first strand and the second

strand form a staggered parallel heterodimer coiled coil structure.

Claim 2 has now in a depending claim the feature of original claim 1 concerning the extension of a first monomer unit in the first strand beyond a corresponding second monomer unit in the second strand in the direction of the strands. In view of the definition of staggered heterodimer at the bottom of page 6 of the description on file (of the published application WO 01/21646) it is considered that claim 2 is superfluous.

7. Claim 1 is considered to miss an essential technical feature of the invention (Article 6 PCT): as mentioned above, the residue other than the normal hydrophobic residue in the "a" position should only be specified as Asn. In addition, it is not clear what the meaning is of "other complementary residues".

8. Claim 3 refers also to an essential technical feature of the invention. Coiled coils are protein-folding motifs that direct and cement a wide variety of protein-protein interactions (see present description page 1, line 10). The present examples also refer all to coiled coils made up of monomer units comprising a **charged** heptad repeat motifs. The feature of claim 3 is therefore essential to the invention and should be incorporated into **claim 1** (to form a staggered, parallel heterodimer). Article 6 of the PCT requires that all independent claims contain the essential technical feature(s) of the invention (see also Rule 6.3(b) PCT).

9. It appears furthermore that units SAF-p1 and SAF-p2C are identical; claim 15 is therefore not concise (Article 6 PCT).

****

peptides at 100 μM (a width value of "*x*" on the histogram includes all measurements made from "(*x*–*5*) to *x*").

Fig. 5 is a cartoon showing the possible anti-typic association of parallel helical peptides leading to a homo-oligomeric peptide nanotube.

Fig. 6 is an x-ray diffraction pattern of an aligned protein fibre of the invention.

Fig. 7 is an image from a confocal fluorescent microscope showing fibres which have been derivatised through the inclusion of flurophores; and

Fig. 8 shows amino acid sequences designed to form blunt-ended heterodimers.

1) Peptide Design and Synthesis

Various peptide monomer units were designed as described above. The monomers and capping peptides (designed to complement the sticky ends of the monomers so as to produce flush, or blunt ends and, so, arrest longitudinal fibre assembly) are set out in Table 1:

principle of using "sticky ends" is well developed in molecular biology for assembling DNA (S. J. Palmer *et al* (1998) *Nucleic Acids Res.* 26, 2560), and has been used to design intricate DNA crystals (E. Winfree *et al* (1998) *Nature* 394, 539). However, to our knowledge, our application of sticky end-directed molecular assembly to peptides is new; although we do note that head-to-tail packing of helices has been observed in recently solved crystal structures for two designer peptides (N. L. Ogihara *et al* (1997) *Protein Sci.* 6, 80; G. G. Prive *et al* (1999) *Protein Sci.* 8, 1400). These were helical peptides that crystallised with their helical ends in contact so as to form pseudo-continuous helices in the solid state. In other words they formed "blunt-ended" arrangements.

US-A-5,712, 366 discloses self-assembling protein material but does not provide details of how to make a staggered parallel heterodimer. WO 96/11947 discloses protein nanostructures based on bacteriophage T4 tail fiber proteins but does not disclose a staggered parallel heterodimer coiled coil structure.

Pandya et al., Biochemistry, 29, 8728-34, 2000 (published after the priority date of the present application) does not disclose a method of making nanotubes and does not disclose a matrix comprising the protein structures of the present invention.

According to one aspect of the invention there is provided a protein structure comprising a plurality of first peptide monomer units arranged in a first strand and a plurality of second peptide monomer units arranged in a second strand, the strands preferably forming a coiled-coil structure, and in which a first peptide monomer unit in the first strand extends beyond a corresponding second peptide monomer unit in the second strand in the direction of the strands. The protein structures of the invention have numerous advantages. For example, relatively long protein fibres can be formed with little material - 1 µl of a 100 µM solution of the peptide monomers may provide enough material to form 10 m of fibre 50 nm thick.

At least one charged amino acid residue of the first peptide monomer unit may be arranged to attract an oppositely-charged amino acid residue of the second peptide monomer unit. Preferably, the charged amino acid residue is in an end portion of the first peptide monomer unit, which extends beyond the corresponding second peptide monomer unit in the second strand. At least one strand may consist solely of first or second peptide monomer units

respectively i.e homogenous strands. Heterologous strands are also contemplated. The peptide monomer units may comprise a repeating structural unit. Preferably, the repeating structural unit comprises a heptad repeat motif, having the pattern:

```
hpphppp
abcdefg
```

Preferably, the repeat may include isoleucine or asparagine at position a and leucine at position d. Other repeats (e.g hendecads - abcdefghijk) and amino acid compositions may also be used (see WO99/11774).

VERSION

PCT

LD INTELLECTUAL PROPERTY ORGANIZAT
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C07K 14/195, C12P 21/06, C07H 17/00

A1

(11) International Publication Number: WO 96/11947

(43) International Publication Date: 25 April 1996 (25.04.96)

(21) International Application Number: PCT/US95/13023

(22) International Filing Date: 13 October 1995 (13.10.95)

(30) Priority Data:
322,760 13 October 1994 (13.10.94) US

(71)(72) Applicant and Inventor: GOLDBERG, Edward, B. [US/US]; 494 Ward Street, Newton, MA 02159-1136 (US).

(74) Agents: MORRIS, Francis, E. et al.; Pennie & Edmonds, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AL, AM, AU, BB, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, IS, JP, KG, KP, KR, KZ, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
With international search report.

(54) Title: MATERIALS FOR THE PRODUCTION OF NANOMETER STRUCTURES AND USE THEREOF

(57) Abstract

The present invention pertains to nanostructures, i.e., nanometer sized structures useful in the construction of microscopic and macroscopic structures. In particular, the present invention pertains to nanostructures based on bacteriophage T4 tail fiber proteins and variants thereof.

* (Referred to in PCT Gazette No. 25/1996, Section II)

SNSDOCID: <WO 9611947A1 IA>

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SI	Slovenia
SK	Slovakia
SN	Senegal
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

ENSDOCID: <WO___9611947A1_IA>

MATERIALS FOR THE PRODUCTION OF NANOMETER STRUCTURES AND USE THEREOF

FIELD OF THE INVENTION

The present invention pertains to nanostructures, *i.e.*, nanometer sized structures useful in the construction of microscopic and macroscopic structures. In particular, the present invention pertains to nanostructures based on bacteriophage T4 tail fiber proteins and variants thereof.

BACKGROUND TO THE INVENTION

While the strength of most metallic and ceramic based materials derives from the theoretical bonding strengths between their component molecules and crystallite surfaces, it is significantly limited by flaws in their crystal or glass-like structures. These flaws are usually inherent in the raw materials themselves or developed during fabrication and are often expanded due to exposure to environmental stresses.

The emerging field of nanotechnology has made the limitations of traditional materials more critical. The ability to design and produce very small structures (*i.e.*, of nanometer dimensions) that can serve complex functions depends upon the use of appropriate materials that can be manipulated in predictable and reproducible ways, and that have the properties required for each novel application.

Biological systems serve as a paradigm for sophisticated nanostructures. Living cells fabricate proteins and combine them into structures that are perfectly formed and can resist damage in their normal environment. In some cases, intricate structures are created by a process of self-assembly, the instructions for which are built into the component polypeptides. Finally, proteins are subject to proofreading processes that insure a high degree of quality control.

Therefore, ther is a n ed in the art for methods and compositions that xploit th s unique features of

SDOCID: <WO 9611947A1 IA>

proteins to form constituents of synthetic nanostructures. The need is to design materials whose properties can be tailored to suit the particular requirements of nanometer-scale technology. Moreover, since the subunits of most macrostructural materials, ceramics, metals, fibers, etc., are based on the bonding of nanostructural subunits, the fabrication of appropriate subunits without flaws and of exact dimensions and uniformity should improve the strength and consistency of the macrostructures because the surfaces are more regular and can interact more closely over an extended area than larger, more heterogeneous material.

SUMMARY OF THE INVENTION

In one aspect, the present invention provides isolated protein building blocks for nanostructures, comprising modified tail fiber proteins of bacteriophage T4. The gp34, 36, and 37 proteins are modified in various ways to form novel rod structures with different properties. Specific internal peptide sequences may be deleted without affecting their ability to form dimers and associate with their natural tail fiber partners. Alternatively, they may be modified so that they: interact only with other modified, and not native, tail fiber partners; exhibit thermolabile interactions with their partners; or contain additional functional groups that enable them to interact with heterologous binding moieties.

The present invention also encompasses fusion proteins that contain sequences from two or more different tail fiber proteins. The gp35 protein, which forms an angle joint, is modified so as to form average angles different from the natural average angle of 137° (±7°) or 156° (±12°), and to exhibit thermolabile interactions with its partners.

In another aspect, the present invention provides nanostructures comprising native and modified tail fiber proteins of bacteriophage T4. The nanostructures may be one-dimensional rods, two-dimensional polygons or open or closed sheets, or three-dimensional open cages or closed solids.

UNSDOCID: <WO__9611947A1_IA>

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A and 1B show a schematic representation of th T4 bacteriophage particle (Figure 1A), and a schematic repr sentation of the T4 bacteriophage tail fiber (Figure 1B).

Figure 2 shows a schematic representation of a unit rod.

Figures 3A-3D show schematic representations of: a one-dimensional multi-unit rod joined along the x axis (Figure 3A); closed simple sheets (Figure 3B); closed brickwork sheets (Figure 3C); and open brickwork sheets (Figure 3D).

Figure 4 shows a schematic representation of two units used to construct porous and solid sheets (top and bottom), which, when alternatively layered, produce a multi-tiered set of cages as shown.

Figure 5 shows a schematic representation of an angled structure having an angle of 120°.

Figure 6 shows the DNA sequence (SEQ ID NO:1) of genes 34, 35, 36, and 37 of bacteriophage T4.

Figure 7 shows the amino acid sequences (shown in single-letter codes) of the gene products of genes 34 (SEQ ID NO:2, ORFX SEQ ID NO:3), 35 (SEQ ID NO:4), 36 (SEQ ID NO:5), and 37 (SEQ ID NO:6) of bacteriophage T4. The amino acid sequences (bottom line of each pair) are aligned with the nucleotide sequences (top line of each pair.) It is noted that the deduced protein sequence of gene 35 (from NCBI database) is not believed to be accurate.

Figures 8A-8B show a schematic representation of: the formation of a P37 dimer initiator from a molecule that self-assembles into a dimer (Figure 8A); and the formation of a P37 trimer initiator from a molecule that self-assembles into a trimer (Figure 8B).

Figure 9 shows a schematic representation of the formation of the polymer (P37-36)n with an initiator that is a self-assembling dimer.

ENSDOCID: <WO 9611947A1 IA>

DETAILED DESCRIPTION OF THE INVENTION

All patents, patent applications and literature references cited in the specification are hereby incorporated by reference in their entirety. In the case of inconsistencies, the present disclosure, including, definitions, will prevail.

Although the invention is described in terms of bacteriophage T4 tail fiber proteins, it will be understood that the invention is also applicable to tail fiber proteins of other T-even-like phage, *e.g.*, of the T4 family (*e.g.*, T4, TuIa, TuIb), and T2 family (T2, T6, K3, Ox2, M1, etc.)

DEFINITIONS:

"Nanostructures" are defined herein as structures of different sizes and shapes that are assembled from nanometer- sized protein components.

"Chimers" are defined herein as chimeric proteins in which at least the amino- and carboxy-terminal regions are derived from different original polypeptides, whether the original polypeptides are naturally occurring or have been modified by mutagenesis.

"Homodimers" are defined herein as assemblies of two substantially identical protein subunits that form a defined three-dimensional structure.

The designation "gp" denotes a monomeric polypeptide, while the designation "P" denotes homooligomers. P34, P36, and P37 are presumably homodimers or homotrimers.

An isolated polypeptide that "consists essentially of" a specified amino acid sequence is defined herein as a polypeptide having the specified sequence or a polypeptide that contains conservative substitutions within that sequence. Conservative substitutions, as those of ordinary skill in the art would understand, are ones in which an acidic residue is replaced by an acidic residue, a basic residue by a basic residue, or a hydrophobic residue by a hydrophobic residue. Also encompassed is a polypeptide that lacks one or more amino acids at either the amino terminus or

BNSDOCID: <WO___9611947A1_IA>

carboxy terminus, up to a total of five at either terminus, when the absence of the particular residues has no discernable effect on the structure or the function of the polypeptide in practicing the present invention.

The present invention pertains to a new class of protein building blocks whose dimensions are measured in nanometers, which are useful in the construction of microscopic and macroscopic structures. Without wishing to be bound by theory, it is believed that the basic unit is a homodimer composed of two identical protein subunits having a cross-β configuration, although a trimeric structure is also possible. Thus, as will be apparent, references to a "homodimer" or "dimerization" as used herein will in many instances be construed as also referring to a homotrimer or trimerization. These long, stiff, and stable rod-shaped units can assemble with other rods using coupling devices that can be attached genetically or *in vitro*. The ends of one rod may attach to different ends of other rods or similar rods. Variations in the length of the rods, in the angles of attachment, and in their flexibility characteristics permit differently-shaped structures to self-assemble *in situ*. In this manner the units can self-assemble into predetermined larger structures of one, two or three dimensions. The self-assembly can be staged to form structures of precise dimensions and uniform strength due to the flawless biological manufacture of the components. The rods can also be modified by genetic and chemical modifications to form predetermined specific attachment sites for other chemical entities, allowing the formation of complex structures.

An important aspect of the present invention is that the protein units can be designed so that they comprise rods of different lengths, and can be further modified to include features that alter their surface properties in predetermined ways and/or influence their ability to join with other identical or different units. Furthermore, the self-assembly capabilities can be expanded by producing chimeric proteins that combine the properties of two

ENSDOCID: <WO 9611947A1 IA-

different members of this class. This design feature is achieved by manipulating the structure of the genes encoding these proteins.

As detailed below, the compositions and methods of the present invention take advantage of the properties of th natural proteins, *i.e.*, the resulting structures are stiff, strong, stable in aqueous media, heat resistant, protease resistant, and can be rendered biodegradable. A large quantity of units can be fabricated easily in microorganisms. Furthermore, for ease of automation, large quantities of parts and subassemblies can be stored and used as needed.

The sequences of the protein subunits are based on the components of the tail fiber of the T4 bacteriophage of *E. coli*. It will be understood that the principles and techniques can be applied to the tail fibers of other T-even phages, or other related bacteriophages that have similar tail and/or fiber structures.

The structure of the T4 bacteriophage tail fiber (illustrated in Figure 1) can be represented schematically as follows (N= amino terminus, C= carboxy terminus): N[P34]C - N[gp35]C - N[P36]C - N[P37]C. P34, P36, and P37 are all stiff, rod-shaped protein homodimers in which two identical β sheets, oriented in the same direction, are fused face-to-face by hydrophobic interactions between the sheets juxtaposed with a 180° rotational axis of symmetry through the long axis of the rod. (The structure will vary if P34, P36, and P37 are homotrimers.) gp35, by contrast, is a monomeric polypeptide that attaches specifically to the N-terminus of P36 and then to the C-terminus of P34 and forms an angle joint between two rods. During T4 infection of *E. coli*, two gp37 monomers dimerize to form a P37 homodimer; the process of dimerization is believed to initiate near the C-terminus of P37 and to require two *E. coli* chaperon proteins. (A variant gp37 with a temperature sensitive mutation near the C-terminus used in the present invention requires only one chaperon, gp57, for dimerization.) Once dimerized, the N-terminus of P37 initiates the dimerization

3NSDOCID: <WO__9611947A1_IA>

of two gp36 monomers to a P36 rod. The joint between the C-terminus of P36 and the N-terminus of P37 is tight and stiff but noncovalent. The N-terminus of P36 then attaches to a gp35 monomer; this interaction stabilizes P36 and forms the elbow of the tail fiber. Finally, gp35 attaches to the C-terminus of P34 (which uses gp57 for dimerization). Thus, self assembly of the tail fiber is regulated by a predetermined order of interaction of specific subunits whereby structural maturation caused by formation of the first subassembly permits interaction with new (previously disallowed) subunits. This results in the production of a structure of exact specifications from a random mixture of the components.

In accordance with the present invention, the genes encoding these proteins may be modified so as to make rods of different lengths with different combinations of ends. The properties of the native proteins are particularly advantageous in this regard. First, the β-sheet is composed of antiparallel β-strands with β-bends at the left (L) and right (R) edges. Second, the amino acid side chains alternate up and down out of the plane of the sheet. The first property allows bends to be extended to form symmetric and specific attachment sites between the L and R surfaces, as well as to form attachment sites for other structures. In addition, the core sections of the β-sheet can be shortened or lengthened by genetic manipulations *e.g.*, by splicing DNA regions encoding β-bends, on the same edge of the sheet, to form new bends that exclude intervening peptides, or by inserting segments of peptide in an analogous manner by splicing at bend angles. The second property allows amino acid side chains extending above and below the surface of the β-sheet to be modified by genetic substitution or chemical coupling. Importantly, all of the above modifications are achieved without compromising the structural integrity of the rod. It will be understood by one skilled in the art that these properties allow a great deal of flexibility in

SDOCID: <WO__9611947A1 IA>

designing units that can assemble into a broad vari ty of structures, som of which are d tail d below.

STRUCTURAL UNITS

The rods of the present invention function like wooden 2 X 4 studs or steel beams for construction. In this case, the surfaces are exactly reproducible at the molecular level and thereby fitted for specific attachments to similar or different units rods at fixed joining sites. The surfaces are also modified to be more or less hydrophilic, including positively or negatively charged groups, and have protrusions built in for specific binding to other units or to an intermediate joint with two receptor sites. The surfaces of the rod and a schematic of the unit rod are illustrated in Figure 2. The three dimensions of the rod are defined as: x, for the back (B) to front (F) dimension; y, for the down (D) to up (U) dimension; and z, for the left (L) to right (R) dimension.

One dimensional multi-unit rods can be most readily assembled from single unit rods joined along the x axis (Figure 3A) but regular joining of subunits in either of th other two dimensions will also form a long structure, but with different cross sections than in the x dimension.

Two dimensional constructs are sheets formed by interaction of rods along any two axes. 1) Closed simple sheets are formed from surfaces which overlap exactly, along any two axes (Figure 3B). 2) Closed brickwork sheets are formed from interaction between units that have exactly overlapping surfaces in one dimension and a special type of overlap in the other (Figure 3C). In this case there must b two different sets of complementary joints spaced with exactly 1/2 unit distance between them. If they are center d (*i.e.*, each set 1/4 from the end) then each joint will be in the center of the units above and below. If they are offset, then the joint will be offs t as well. In this construction, the complementary interacting ites are schematized by • and ••. If th int racting sit s are ach symmetric, the

ISDOCID: <WO___9611947A1_IA>

alt rnating rows can interact with the rods in either direction. If they are not symmetric, and can nly interact with interacting r ws facing in the same r opposit direction, the sheet will made of unidirectional rods or layers of rods in alternating directions. 3) Open brickwork sheets (or nets) result when the units are separated by more than one-half unit (Figure 3D). The dimensions of the openings (or pores) depend upon the distance (dx) separating the interacting sites and the distance (dy) by which these sites separate the surfaces.

Three dimensional constructs require sterically compatible interactions between all three surfaces to form solids. 1) Closed solids can assemble from units that overlap exactly in all three dimensions (*e.g.*, the exact overlapping of closed simple sheets). In an analogous manner, closed brickwork sheets can form closed solids by overlapping sheets exactly or displaced to bring the brickwork into the third dimension. This requires an appropriate set of joints on all three pairs of parallel faces of the unit. 2) Porous solids are made by joining open brickwork sheets in various ways. For example, if the units overlap exactly in the third dimension, a solid is formed with the array of holes of exact dimensions running perpendicular to the plane of the paper. If instead, a material is needed with closed spaces, with layers of width dz (*i.e.*, in the U→>D dimension), a simple closed sheet is layered on the open brickwork sheet to close the openings. If the overlap of the open brickwork sheet is *e.g.*, 1/4 unit, then a rod of length 3/4 units is used to make the sheet. Joints are then needed in the z dimension. The two units used to polymerize these alternate layers, and the layers themselves, are schematized in Figure 4.

All of the above structures are composed of simple linear rods. A second unit, the angle unit, expands the type and dimensi nality of possible structures. The angle unit connects two rods at angles diff rent from 180°, akin to an angle ir n. The average angle and its degree of rigidity are

built int this conn ctor structure. F r exampl , the structure shown in Figure 5 has an angle of 120 and different sp cific j ining sites at a and at b. Th following are exampl s of structures that are form d utilizing angle joints:

1) Open brickwork sheets are expanded and strengthened in the direction normal to the rod direction by adding angles perpendicular to the sheet. In this case, a three dimensional network forms. Attachment of 90° angles to the ends of the rods makes an angle almost in the plane of the sheet, allowing new rods added to those angles (which must have some play out of the plane of the original sheet to attach in the first place) to form a new sheet, almost parallel, with an orientation normal to its upper or lower neighbor.

2) Hexagons are made from a mixture of rods and angle joints that form 120° angles. In this case, there are two exclusive sets of joints. Each set is made up of one f the two ends of the rod and one of the two complementary sites on the angle. This is a linear structure in the sense that the hexagon has a direction (either clockwise or counterclockwise). It can be made into a two dimensional open net (*i.e.*, a two dimensional honeycomb) by joining the sides of the hexagons. It can form hexagonal tubes by joining the top of the hexagon below to the bottom face of the hexagon above. If the tubes also join by their sides, they will form an open three dimensional multiple hexagonal tube.

3) Helical hexagonal tubes are made analogously to hexagons but the sixth unit is not joined to the first to close the hexagon. Instead, the end is displaced from the plane of the hexagon and the seventh and further units are added to form a hexagonal tube which can be a spring if there is little or no adhesive force between the units of the helix, or a stiff rod if there is such a force to maintain the close pr ximity of app sing units.

BNSDOCID: <WO__9611947A1_IA>

It will be apparent to one skilled in the art that the compositions and methods of the present invention also encompass other polygonal structures such as octagons, as well as open solids such as tetrahedrons and icosahedrons formed from triangles and boxes formed from squares and rectangles. The range of structures is limited only by the types of angle units and the substituents that can be engineered on the different axes of the rod units. For example, other naturally occurring angles are found in the fibers of bacteriophage T7, which has a 90° angle (Steven et al., *J. Mol. Biol. 200*: 352-365, 1988).

DESIGN AND PRODUCTION OF THE ROD PROTEINS

The protein subunits that are used to construct the nanostructures of the present invention are based on the four polypeptides that comprise the tail fibers of bacteriophage T4, *i.e.*, gp34, gp35, gp36 and gp37. The genes encoding these proteins have been cloned, and their DNA and protein sequences have been determined (for gene 36 and 37 see Oliver et al. *J. Mol. Biol. 153*: 545-568, 1981). The DNA and amino acid sequences of genes 34, 35, 36 and 37 are set forth in Figures 6 and 7 below.

Gp34, gp35, gp36, and gp37 are produced naturally following infection of *E. coli* cells by intact T4 phage particles. Following synthesis in the cytoplasm of the bacterial cell, the gp34, 36, and 37 monomers form homodimers, which are competent for assembly into maturing phage particles. Thus, *E. coli* serves as an efficient and convenient factory for synthesis and dimerization of the protein subunits described herein below.

In practicing the present invention, the genes encoding the proteins of interest (native, modified, or recombined) are incorporated into DNA expression vectors that are well known in the art. These circular plasmids typically contain selectable marker genes (usually conferring antibiotic resistance to transformed bacteria), sequences that allow replication of the plasmid to high copy number in

E. coli, and a multiple cloning site immediately d wnstream of an inducible promoter and ribosome binding site. Examples f c mmercially available v ctors suitable f r use in the present invention include the pET system (Novagen, Inc., Madison, WI) and Superlinker vectors pSE280 and pSE380 (Invitrogen, San Diego, CA).

The strategy is to 1) construct the gene of interest and clone it into the multiple cloning site; 2) transform *E. coli* cells with the recombinant plasmid; 3) induce the expression of the cloned gene; 4) test for synthesis of the protein product; and, finally, 5) test for the formation of functional homodimers. In some cases, additional genes are also cloned into the same plasmid, when their function is required for dimerization of the protein of interest. For example, when wild-type or modified versions of gp37 are expressed, the bacterial chaperon gene 57 is also included; when wild-type or modified gp36 is expressed, the wild-type version or a modified version of the gp37 gene is included. The modified gp37 should have the capacity to dimerize and contain an N-terminus that can chaperon the dimerization of gp36. This method allows the formation of monomeric gene products and, in some cases, maturation of monomers to homodimeric rods in the absence of other phage-induced proteins normally present in a T4-infected cell.

Steps 1-4 of the above-defined strategy are achieved by methods that are well known in the art of recombinant DNA technology and protein expression in bacteria. For example, in step 1, restriction enzyme cleavage at multiple sites, followed by ligation of fragments, is used to construct deletions in the internal rod segment of gp34, 36, and 37 (see Example 1 below). Alternatively, a single or multiple restriction enzyme cleavage, followed by exonuclease digestion (EXO-SIZE, New England Biolabs, Beverly, MA), is used to d lete DNA sequences in one or both directions from th initial cleavage site; when combined with a subsequent ligation step, this

3NSDOCID: <WO___9611947A1_IA>

procedure produces a nested s t f deletions of increasing sizes. Similarly, standard m thods are used to r combine DNA segments from two different tail fiber genes, to produc chimeric genes encoding fusion proteins (called "chimers" in this description). In general, this last method is used to provide alternate N- or C-termini and thus create novel combinations of ends that enable new patterns of joining of different rod segments. A representative of this type of chimer, the fusion of gp37-36, is described in Example 2. The preferred hosts for production of these proteins (Step 2) is *E. coli* strain BL21(DE3) and BL21(DE3/pLysS) (available commercially from Novagen, Madison, WI), although other compatible recA strains, such as HMS174(DE3) and HMS174(DE3/pLysS) can be used. Transformation with the recombinant plasmid (Step 2) is accomplished by standard methods (Sambrook, J., *Molecular Cloning*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY; this is also the sourc for standard recombinant DNA methods used in this invention.) Transformed bacteria are selected by virtue of their resistance to antibiotics *e.g.*, ampicillin or kanamycin. The method by which expression of the cloned tail fiber genes is induced (Step 3) depends upon the particular promoter used. A preferred promoter is plac (with a $laci^q$ on the vector to reduce background expression), which can be regulated by the addition of isopropylthiogalactoside (IPTG). A second preferred promoter is pT7ϕ10, which is specific to T7 RNA polymerase and is not recognized by *E. coli* RNA polymerase. T7 RNA polymerase, which is resistant to rifamycin, is encoded on the defective lambda DE lysogen in the *E. coli* BL21 chromosome. T7 polymerase in BL21(DE3) is super-repressed by the $laci^q$ gene in the plasmid and is induced and regulated by IPTG.

Typically, a culture of transformed bacteria is incubated with the inducer for a period of hours, during which th synthesis of the prot in of interest is monitored. In the present instance, extracts of the bacterial cells are

SNSDOCID: <WO 9611947A1 1>

prepared, and the T4 tail fiber proteins are detected, for example, by SDS-polyacrylamide gel electrophoresis.

Once the modified protein is detected in bacterial extracts, it is necessary to ascertain whether or not it forms appropriate homodimers (Step 4). This is accomplished initially by testing whether the protein is recognized by an antiserum specific to the mature dimerized form of the protein.

Tail fiber-specific antisera are prepared as described (Edgar, R.S. and Lielausis, I., *Genetics* 52: 1187, 1965; Ward et al, *J. Mol. Biol.* 54:15, 1970). Briefly, whole T4 phage are used as an immunogen; optionally, the resulting antiserum is then adsorbed with tail-less phage particles, thus removing all antibodies except those directed against the tail fiber proteins. In a subsequent step, different aliquots of the antiserum are adsorbed individually with extracts that each lack a particular tail fiber protein. For example, if an extract containing only tail fiber components P34, gp35, and gp36 (derived from a cell infected with a mutant T4 lacking a functional gp37 gene) is used for absorption, the resulting antiserum will recognize only mature P37 and dimerized P36-P37. A similar approach may be used to prepare individual antisera that recognize only mature (*i.e.*, homodimerized) P34 and P36 by adsorbing with extracts containing distal half tail fibers or P34, gp35 and P37, respectively. An alternative is to raise antibody against purified tail fiber halves, *e.g.*, P34 and gp35-P36-P37. Anti gp35-P36-P37 can then be adsorbed with P36-P37 to produce anti-gp35, and anti-P36 can be produced by adsorption with P37 and gp35. Anti-P37, anti-gp35, and anti-P34 can also be produced directly by using purified P37, gp35, and P34 as immunogens. Another approach is to raise specific monoclonal antibodies against the different tail fiber components or segments thereof.

Specific antibodies to subunits or tail parts are used in any of the following ways to detect appropriately homodimerized tail fiber proteins: 1) Bacterial colonies are

INSDOCID: <WO___9611947A1_IA>

screened for those expressing mature tail fiber proteins by directly transferring the colonies, or, alternatively, samples of lysed or unlysed cultures, to nitrocellulose filters, lysing the bacterial cells on the filter if necessary, and incubating with specific antibodies. Formation of immune complexes is then detected by methods widely used in the art (*e.g.*, secondary antibody conjugated to a chromogenic enzyme or radiolabelled Staphylococcal Protein A.). This method is particularly useful to screen large numbers of colonies *e.g.*, those produced by EXO-SIZE deletion as described above. 2) Bacterial cells expressing the protein of interest are first metabolically labelled with ^{35}S-methionine, followed by preparation of extracts and incubation with the antiserum. The immune complexes are then recovered by incubation with immobilized Protein A followed by centrifugation, after which they may be resolved by SDS-polyacrylamide gel electrophoresis.

An alternative competitive assay for testing whether internally deleted tail fiber proteins that do not permit phage infection nonetheless retain the ability to dimerize and associate with their appropriate partners utilizes an *in vitro*, complementation system. 1) A bacterial extract containing the modified protein of interest, as described above, is mixed with a second extract prepared from cells infected with a T4 phage that is mutant in the gene of interest. 2) After several hours of incubation, a third extract is added that contains the wild-type version of the protein being tested, and incubation is continued for several additional hours. 3) Finally, the extract is titered for infectious phage particles by infecting *E. coli* and quantifying the phage plaques that result. A modified tail fiber protein that is correctly dimerized and able to join with its partners is incorporated into tail fibers in a non-functional manner in Step 1, thereby preventing the incorporation of the wild-type version of the protein in Step 2; the result is a <u>reduction</u> in the titer of the resulting phage sample. By contrast, if the modified protein is unable

SDOCID: <WO___9611947A1_IA>

t dimerize and thus form proper N- and/or C-termini, it will not be inc rporated into phage particl s in St p 1, and thus will n t compet with assembly f intact phag particl s in Step 2; the phage titer sh uld thus be equivalent to that observed when no modified protein is added in Step 1 (a negative control.)

Another way in which to test whether chimers and internally deleted tail fiber proteins retain the ability to dimerize and associate with their appropriate partners is done *in vivo*. The assay detects the ability of such chimers and deleted proteins to compete with normal phage parts for assembly, thus reducing the burst size of a wild-type phag infecting the same host cell in which the chimers or deleted proteins are recombinantly expressed. Thus, expression from an expression vector encoding the chimer or deleted protein is induced inside a cell, which cell is then infected by a wild-type phage. Inhibition of wild-type phage production demonstrates the ability of the recombinant chimer or protein to associate with the appropriate tail fiber proteins of the phage.

The above-described methods are used, alone and in combination, in the design and production of different types of modified tail fiber proteins. For example, a preliminary screen of a large number of bacterial colonies for those expressing a properly dimerized protein will identify positive colonies, which can then be individually tested by *in vitro* complementation.

Non-limiting examples of novel proteins that are encompassed by the present invention include:

1) Internally deleted gp34, 36, and 37 polypeptides (See Example 1 below);

2) A C-terminally truncated gp36 fused to the N-terminus of N-terminally truncated gp37;

3) A fusion between gp36 and gp37 in which gp37 is N-terminal to gp36 (*i.e.*, in reverse of the natural rder), termed her in "gp37-36 chimer" (See Example 2 bel w);

NSDOCID: <WO___9611947A1_IA>

4) A fusion between gp34 and gp36 in which gp36 is N-terminal to gp34 (*i.e.*, in reverse of the natural order), termed herein "gp36-34 chimer";

5) A variant of gp36 in which the C-terminus is mutated such that it lacks the capability to interact with (and dimerize in response to) the N-terminus of wild-type P37, termed herein "gp36*";

6) A variant of gp37 in which the N-terminus is mutated such that it forms a P37 that lacks the capability to interact with the C-terminus of wild-type gp36, termed herein "*P37";

7) Variants of gp36* and *P37 that can interact with each other, but not with gp36 or P37.

8) A variant "P37-36 chimer" in which the gp36 moiety is derived from the variant as in 5), *i.e.*, "P37-36*". (For 5-8, See Example 3 below.)

9) A variant "P37-36 chimer" in which the gp37 moiety is derived from the variant as in 6) above, *i.e.*, "*P37-36".

10) A variant P37-36 chimer, *P37-P36*, in which the gp36 and gp37 moieties are derived from the variants in 7).

11) A fusion between gp36 and gp34 in which gp36 sequences are placed N-terminal to gp34, the dimer of which is termed herein "P36-34 chimer";

12) Variants of gp35 that form average angles different from 137° or 158° (the native angle) *e.g.*, less than about 125° or more than about 145° under conditions wherein the wild-type gp35 protein forms an angle of 137° when combined with the P34 and P36-P37 dimers, and/or exhibit more or less flexibility than the native polypeptide;

13) Variants of gp34, 35, 36 and 37 that exhibit thermolabile interactions or other variant specific interactions with their cognate partn rs; and

14) Variants of gp37 in which the C-terminal domain f the polypeptide is modified to include sequences that c nfer specific binding properties on the entir

NSDOCID: <WO___9611947A1 IA>

molecules, e.g., sequences derived from avidin that recognize biotin, sequences derived from immunoglobulin heavy chain that recognize Staphylococcal A protein, sequences derived from the Fab portion of the heavy chain of monoclonal antibodies to which their respective Fab light chain counterparts could attach and form an antigen-binding site, immunoactive sequences that recognize specific antibodies, or sequences that bind specific metal ions. These ligands may be immobilized to facilitate purification and/or assembly.

In specific embodiments, the chimers of the invention comprise a portion consisting of at least the first 10 (N-terminal) amino acids of a first tail fiber protein fused via a peptide bond to a portion consisting of at least the last 10 (C-terminal) amino acids of a second tail fiber protein. The first and second tail fiber proteins can be the same or different proteins. In another embodiment, the chimers comprise an amino acid portion in the range of the first 10-60 amino acids from a tail fiber protein fused to an amino acid portion in the range of the last 10-60 amino acids from a second tail fiber protein. In another embodiment, each amino acid portion is at least 20 amino acids of the tail fiber protein. The chimers comprise portions, *i.e.*, not full-length tail fiber proteins, fused to one another. In a preferred aspect, the first tail fiber protein portion of the chimer is from gp37, and the second tail fiber protein portion is from gp36. Such a chimer (gp37-36 chimer), after oligomerization to form P37-36, can polymerize to other identical oligomers. A gp36-34 chimer, after oligomerization to form P36-34, can bind to gp35, and this unit can then polymerize. In another embodiment, the first portion is from gp37, and the second portion is from gp34. In a preferred aspect, the chimers of the invention are made by insertions or deletions within a β turn of the β structure of the tail fiber proteins. Most preferably, insertions into a tail fiber sequence, or fusing to another tail fiber protein sequence, (preferably via manipulation at the recombinant DNA level to produce the desired encoded protein) is done so that

NSDOCID: <WO___9611947A1_IA>

sequences in β turns on the same edge of the β-sheet are joined.

In addition to the above-described chimers, nanostructures of the invention can also comprise tail fiber protein deletion constructs that are truncated at one end, *e.g.*, are lacking an amino- or carboxy- end (of at least 5 or 10 amino acids) of the molecule. Such molecules truncated at the amino-terminus, *e.g.*, of truncated gp37, gp34, or gp36, can be used to "cap" a nanostructure, since, once incorporated, they will terminate polymerization. Such molecules preferably comprise a fragment of a tail fiber protein lacking at least the first 10, 20, or 60 amino terminal amino acids.

In order to change the length of the rod component proteins as desired, portions of the same or different tail fiber proteins can be inserted into a tail fiber chimer to lengthen the rod, or be deleted from a chimer, to shorten the rod.

ASSEMBLY OF INDIVIDUAL ROD COMPONENTS INTO NANOSTRUCTURES

Expression of the proteins of the present invention in *E. coli* as described above results in the synthesis of large quantities of protein, and allows the simultaneous expression and assembly of different components in the same cells. The methods for scale-up of recombinant protein production are straightforward and widely known in the art, and many standard protocols can be used to recover native and modified tail fiber proteins from a bacterial culture.

In a preferred embodiment, native (nonrecombinant) gp35 is isolated for use by growing up a bacteriophage T4 having an amber mutation in gene 36, in a su° bacterial strain (not an amber suppressor), and isolating gp35 from the resulting culture by standard methods.

P34, P36-P37, P37, and chimers derived from them are purified from *E. coli* cultures as mature dimers. Gp35 and variants thereof are purified as monomers. Purification is achieved by the following procedures or combinations thereof,

ENSDOCID: <WO 9611947A1 I>

using standard methods: 1) chromatography on molecular sieve, ion- xchang , and/or hydrophobic matrices; 2) preparative ultrac ntrifugati n; and 3) affinity chromatography, using as the immobilized ligand specific antibodies or other specific binding moieties. For example, the C-terminal domain of P37 binds to the lipopolysaccharide of *E. coli* B. Other T4-like phages have P37 analogues that bind other cell surface components such as OmpF or TSX protein. Alternatively, if the proteins have been engineered to include heterologous domains that act as ligands or binding sites, the cognate partner is immobilized on a solid matrix and used in affinity purification. For example, such a heterologous domain can be biotin, which binds to a streptavidin-coated solid phase.

Alternatively, several components are co-expressed in the same bacterial cells, and sub-assemblies of larger nanostructures are purified subsequent to limited *in vivo* assembly, using the methods enumerated above.

The purified components are then combined *in vitro* under conditions where assembly of the desired nanostructure occurs at temperatures between about 4°C and about 37°C, and at pHs between about 5 and about 9. For a given nanostructure, optimal conditions for assembly (*i.e.*, type and concentration of salts and metal ions) are easily determined by routine experimentation, such as by changing each variable individually and monitoring formation of the appropriate products.

Alternatively, one or more crude bacterial extracts may be prepared, mixed, and assembly reactions allowed to proceed prior to purification.

In some cases, one or more purified components assemble spontaneously into the desired structure, without the necessity for initiators. In other cases, an initiator is required to nucleate the polymerization of rods or sheets. This offers th advantage of localizing the assembly pr cess (*i.e.*, if the initiator is imm bilized or therwis localized) and of regulating the dimensions of the final

NSDOCID: <WO___9611947A1_IA>

structur . F r exampl , rod c mp nents that contain a functional P36 C-terminus require a functional P37 N-terminus to initiate rod formation st ichi metrically; thus, altering the relative amount of initiator and rod component will influence the average length of rod polymer. If the ratio is n, the average rod will be approximately (P37-36)n--N-terminus P37-P37 C-terminus.

In still other cases, the final nanostructure is composed of two or more components that cannot self-assemble individually but only in combination with each other. In this situation, alternating cycles of assembly can be staged to produce final products of precisely defined structure (see Example 6B below.)

When an immobilized initiator is used, it may be desirable to remove the polymerized unit from the matrix after staged assembly. For this purpose specialized initiators are engineered so that the interaction with the first rod component is rendered reversibly thermolabile (see Example 5 below). In this way, the polymer can be easily separated from the matrix-bound initiator, thereby permitting: 1) easy preparation of stock solutions of uniform parts or subassemblies, and 2) re-use of the matrix-bound initiator for multiple cycles of polymer initiation, growth, and release.

In an embodiment in which a nanostructure is assembled that is attached to a solid matrix via gp34 (or P34), one way in which to detach the nanostructure to bring it into solution is to use a mutant (thermolabile) gp34 that can be made to detach upon exposure to a higher temperature (*e.g.*, 40°C). Such a mutant gp34, termed T4 tsB45, having a mutation at its C-terminal end such that P34 attaches to the distal tail fiber half at 30°C but can be separated from it *in vitro* by incubation at 40°C in the presence of 1% SDS (unlike wild-type T4 which are stable under these conditions), has be n r ported (Seed, 1980, Studies f the Bacteriophage T4 Proximal Half Tail Fiber, Ph.D. Thesis, California Institute of Technology), and can be used.

SDOCID: <WO___9611947A1 IA>

Proteins which catalyze the formation of correct (lowest energy) stable secondary (2°) structure of proteins are called chaperonin proteins. (Often, especially in globular proteins, this stabilization is aided by tertiary structure, *e.g.*, stabilization of β-sheets by their interaction in β-barrels or by interaction with α-helices). Normally chaperonins prevent intrachain or interchain interactions which would produce untoward metastable folding intermediates and prevent or delay proper folding. There are two known accessory proteins, gp57 and gp38, in the morphogenesis of T4 phage tail fibers which are sometimes called chaperonins because they are essential for proper maturation of the protein oligomers but are not present in the final structures.

The usual chaperonin system (*e.g.*, groEL/ES) interact with certain oligopeptide moieties of the gene product to prevent unwanted interactions with oligopeptide moieties elsewhere on the same polypeptide or another peptide. These would form metastable folding intermediates which retard or prevent proper folding of the polypeptide to its native (lower energy) state.

Gp57, probably in conjunction with some membrane protein(s), has the role of juxtaposing (and aligning) and/or initiating the folding of 2 or 3 identical gp37 molecules. The aligned peptides then zip up (while mutually stabilizing their nascent β-structures) to form a beam, without further interaction with gp57. Gp57 acts in T4 assembly not only for oligomerization of gp37 but also for gp34 and gp12.

STRUCTURAL COMPONENTS FOR SELF ASSEMBLY OF BEAMS *IN VITRO*

Alternatively to starting the polymerization of chimers with the use of a preformed chimeric or natural oligomeric unit called an initiator produced *in vivo*, molecules (preferably peptides) that can self-assemble can be produced as fusion proteins, fused to the N- or C-terminus of tail fiber variants of the invention (chimers, deletion/insertion constructs) to align their ends and thus

BNSDOCID: <WO___9611947A1_IA>

to facilitate their subsequent unaided folding into oligomeric, stable β-folded rod-like (beam) units *in vitro*, in the absence of the normally required chaperonin proteins (*e.g.*, gp57) and host cell membrane proteins.

As an illustration, consider the P37 unit as an initiator of gp37-36 oligomerization and polymerization. Normally, proper folding of gp37 to a P37 initiator requires a phage infected cell membrane, and two chaperone proteins, gp38 and gp57. In a preferred embodiment, the need for gp38 can be obviated by use of a mutation, ts3813 (a duplication of 7 residues just downstream of the transition zone of gp37) which suppresses gene 38 (Wood, W.B., F.A. Eiserling and R.A. Crowther, 1994, "Long Tail Fibers: Genes, Proteins, Structure, and Assembly," in Molecular Biology of Bacteriophage T4, (Jim D. Karam, Editor) American Society for Microbiology, Washington, D.C., pp 282-290). If a moiety that self-assembles into a dimer or trimer or other oligomer ("self-assembling moiety") is fused to a C-terminal deletion of gp37 downstream or upstream of the transition region [the transition region is a conserved 17 amino acid residue region in T4-like tail fiber proteins where the structure of the protein narrows to a thin fiber; see Henning et al., 1994, "Receptor recognition by T-even-type coliphages," in Molecular Biology of Bacteriophage T4, Karam (ed.), American Society for Microbiology, Washington, D.C., pp. 291-298; Wood et al., 1994, "Long tail fibers: Genes, proteins, structure, and assembly," in Molecular Biology of Bacteriophage T4, Karam (ed.), American Society for Microbiology, Washington, D.C., pp. 282-290], when it is expressed, the self-assembling moiety will oligomerize in parallel and thus align the fused gp37 peptides, permitting them to fold *in vitro*, in the absence of other chaperonin proteins.

If P37 is a dimer (Figure 8A), the self-assembling moiety can be a self dimerizing peptide such as the leucin zipper, made from residues 250-281 from the yeast transcription factor, GCN4 (E.K. O'Shea, R.Rutkowski and P.S. Kim, Science 243:538, 1989) or the self dimerizing mutant

SDOCID: <WO___9611947A1 IA>

l ucine zipper peptid , pIL in which the a positions are substituted with isol ucin and the d positions with l ucine (Harbury P.B., T. Zhang, P.S. Kim and T. Alper. 1993. A Switch Between Two-, Three-, and F ur-Stranded Coil d C ils in GCN4 Leucine Zipper Mutants. Science, 262:1401-1407). If P37 is a trimer (Figure 8B), the self-assembling moiety can be a self trimerizing mutant leucine zipper peptide, pII in which both the a and d positions are substituted with isoleucine (Harbury P.B., et al. *ibid*). Alternatively, a collagen peptide can be used as the self-assembling moiety, such as that described by Bella et al. (J. Bella, M. Eaton, B. Brodsky and H.M. Berman. 1994. Crystal and Molecular Structure of a Collagen-Like Peptide at 1.9Å Resolution. Science, 226:75-81), which self aligns by an inserted specific non repeating alanine residue near the center.

Self-assembling moieties can be used to make initiators for polymerizations in the absence of the normal initiators. For example, to create an initiator for oligomerization and polymerization of the chimeric monomer, gp37-36, gp37-36-C_2 can be used as illustrated in Figure 9. (C_2 means that a dimer forming peptide is fused to the C-terminus of the gp36 moiety. This is used if the beam is a dimeric structure. Otherwise C_3 -- a trimer forming peptide fused to the C-terminus -- would be used.) Furthermore, use of the *E. coli lac* repressor N-terminus, *e.g.*, which associates as a tetramer, with two coils facing in each direction could join two dimers (or polymers of dimers) end to end, either at their N- or C-termini depending upon which end the self-assembling peptides were placed. They could also join N- to C- termini. In any case, alone, they could only form a dimer, each end of which would be extensible by adding an appropriate chimer monomer (as shown for the simpler case in Figure 9).

In an alternative embodiment, the self-assembling moiety can be fused to th N-termini of the chimer. In a specific embodiment, th self-assembling moiety is fus d to

ISDOCID: <WO__9611947A1_IA>

at least a 10 amino acid portion of a T-even-like tail fiber protein.

A self assembling moiety that assembles into a heteroligomer can also be used. For example, if polymerization between beams is directed by the surface of a dimeric cross-β surface, addition of a heterodimeric unit with one surface which does not promote further polymerization would be very useful to cap the penultimate unit and thus terminate polymerization. If the two types of coiled regions of the self-assembling moiety are much more attractive to each other that to themselves, then all of the dimers will be heterodimers. Such is the case for the N-terminal Jun and Fos leucine zipper regions.

A further advantage to such heterodimeric units is the ability to stage polymerization and thus build one unit (or one surface in a 2D array) at a time. For example, suppose surface A attaches to B but neither attaches to itself ([A<->B] is used to symbolize this type of interaction). Mix A/A and B/B_o (B_o is attached to a matrix for easy purification). This will form B_o/B-A/A. Now wash out A/A and add B/B. The construct is now B_o/B-A/A-B/B. Now add A/A_o. The construct is now B_o/B-A/A-B/B-A/A_o and no more beams can be added. There are of course many other possibilities.

APPLICATIONS

The uses of the nanostructures of the present invention are manifold and include applications that require highly regular, well-defined arrays of fibers, cages, or solids, which may include specific attachment sites that allow them to associate with other materials.

In one embodiment, a three-dimensional hexagonal array of tubes is used as a molecular sieve or filter, providing regular vertical pores of precise diameter for selective separation of particles by size. Such filters can be used for sterilization of solutions (*i.e.*, to remove microorganisms or viruses), or as a series of

SDOCID: <WO___9611947A1 IA>

molecular-weight cut-off filters. In this cas , th protein c mponents of the pores may be modified so as to provide specific surface pr perties (*i*. ., hydrophilicity or hydr ph bicity, ability to bind specific ligands, etc.). Among the advantages of this type of filtration device is th uniformity and linearity of pores and the high pore to matrix ratio.

In another embodiment, long one-dimensional fibers are incorporated, for example, into paper or cement or plastic during manufacture to provide added wet and dry tensile strength.

In still another embodiment, different nanostructure arrays are impregnated into paper and fabric as anti- counterfeiting markers. In this case, a simple color-linked antibody reaction (such as those commercially available in kits) is used to verify the origin of the material. Alternatively, such nanostructure arrays could bind dyes or other substances, either before or after incorporation to color the paper or fabrics or modify their appearance or properties in other ways.

KITS

The invention also provides kits for making nanostructures, comprising in one or more containers the chimers and deletion constructs of the invention. For example, one such kit comprises in one or more containers purified gp35 and purified gp36-34 chimer. Another such kit comprises purified gp37-36 chimer.

The following examples are intended to illustrate the present invention without limiting its scope.

In the examples below, all restriction enzymes, nucleases, ligases, etc. are commercially available from numerous commercial sources, such as New England Biolabs (NEB), Beverly, MA; Life Technologies (GIBCO-BRL), Gaithersburg, MD; and Boehringer Mannheim Corp. (BMC), Indianapolis, IN.

BNSDOCID: <WO___9611947A1_IA>

EXAMPLE 1

DESIGN, CONSTRUCTION AND EXPRESSION OF INTERNALLY DELETED P37

The gene encoding gp37 contains two sites for the restriction enzyme Bgl II, the first cleavage occurring after nucleotide 293 and the second after nucleotide 1486 (the nucleotides are numbered from the initiator methionine codon ATG.) Thus, digestion of a DNA fragment encoding gp37 with BglII, excision of the intervening fragment (nucleotides 294- 1485) and re-ligation of the 5′ and 3′ fragments results in the formation of an internally deleted gp37, designated ΔP37, in which arginine-98 is joined with serine-497.

The restriction digestion reaction mix contains:

gp37 plasmid DNA (1 μg/μl)	2μl
NEB buffer #2 (10X)	1μl
H_2O	6μl
Bgl II (10 U/μl)	1μl

The gp37 plasmid signifies a pT7-5 plasmid into which gene 37 has been inserted in the multiple cloning site, downstream of a good ribosome binding site and of gene 57 to chaperon the dimerization. The reaction is incubated for 1h at 37°C. Then, 89 μl of T4 DNA ligase buffer and 1 μl of T4 DNA ligase are added, and the reaction is continued at 16°C for 4 hours. 2 μl of the Stu I restriction enzyme are then added, and incubation continued at 37°C for 1h. (The Stu I restriction enzyme digests residual plasmids that were not cut by Bgl II in the first step, reducing their transformability by about 100-fold.)

The reaction mixture is then transformed into *E. coli* strain BL21, obtained from Novagen, using standard procedures. The transformation mixture is plated onto nutrient agar containing 100 μg/ml ampicillin, and the plates are incubated overnight at 37°C.

Colonies that appear after overnight incubation are picked, and plasmid DNA is extracted and digested with Bgl II as above. The restriction digests are resolved on 1% agarose

NSDOCID <WO 9611947A1

gels. A successful deletion is evidenced by the appearanc after gel electroph r sis f a new DNA fragment f 4.2 kbp, repr senting th und leted part f gen 37 which is still attached to the plasmid and which re-f rm d a BglII site by ligation. The 1.2 kbp DNA fragment bounded by BglII sites in the original gene is no longer in the plasmid and so is missing from the gel.

Plasmids selected for the predicted deletion as above are transformed into *E. coli* strain BL21(DE3). Transformants are grown at 30°C until the density (A_{600}) of the culture reaches 0.6. IPTG is then added to a final concentration of 0.4 mM and incubation is continued at 30°C for 2h, after which the cultures are chilled on ice. 20 µl of the culture is then removed and added to 20 µl of a two-fold concentrated "cracking buffer" containing 1% sodium dodecyl sulfate, glycerol, and tracking dye. 15 µl of this solution are loaded onto a 10% polyacrylamide gel; a second aliquot of 15 µl is first incubated in a boiling water bath for 3 min and then loaded on the same gel. After electrophoresis, the gel is fixed and stained. Expression of the deleted gp37 is evidenced by the appearance of a protein species migrating at an apparent molecular mass of 65-70,000 daltons in the boiled sample. The extent of dimerization is suggested by the intensity of higher-molecular mass species in the unboiled sample and/or by the disappearance of the 65-70,000 dalton protein band.

The ability of the deleted polypeptide to dimerize appropriately is directly evaluated by testing its ability to be recognized by an anti-P37 antiserum that reacts only with mature P37 dimers, using a standard protein immunoblotting procedure.

An alternative assay for functional dimerization f the deleted P37 polypeptide (also referred to as ΔP37) is its ability to complement *in vivo* a T4 37⁻ phage, by first inducing xpression of the ΔP37 and then infecting with th T4 mutant, and detecting pr geny phage.

NSDOCID: <WO___9611947A1_IA>

A ΔP37 was prepared as described above, and found capable of complementing a T4 37⁻ phage *in vivo*.

EXAMPLE 2

DESIGN, CONSTRUCTION AND EXPRESSION OF A gp37-36 CHIMER

The starting plasmid for this construction is one in which the gene encoding gp37 is cloned immediately upstream (*i.e.*, 5′) of the gene encoding gp36. The plasmid is digested with Hae III, which deletes the entire 3′ region of gp37 DNA downstream of nucleotide 724 to the 3′ terminus, and also removes the 5′ end of gp36 DNA from the 5′ terminus to nucleotide 349. The reaction mixture is identical to that described in Example 1, except that a different plasmid DNA is used, and the enzyme is HaeIII. Ligation using T4 DNA ligase, bacterial transformation, and restriction analysis are also performed as in Example 1. In this case, excision of the central portion of the gene 37-36 insert and religation reveals a novel insert of 346 in-frame codons, which is cut only once by HaeIII (after nucleotide 725). The resulting construct is then expressed in *E. coli* BL21(DE3) as described in Example 1.

Successful expression of the gp37-36 chimer is evidenced by the appearance of a protein product of about 35,000 daltons. This protein will have the first 242 N-terminal amino acids of gp37 fused to the final 104 C-terminal amino acids of gp36 (numbered 118-221.) The utility of this chimer depends upon its ability to dimerize and attach end-to-end. That is, carboxy termini of said polypeptide will have the capability of interacting with the amino terminus of the P37 protein dimer of bacteriophage T4 and to form an attached dimer, and the amino terminus of the dimer of said polypeptide will have the capability of interacting with other said chimer polypeptides. This property can be tested by assaying whether introduction of ΔP37 initiates dimerization and polymerization. Alternatively, polyclonal antibodies specific to P36 dimer

BNSDOCID: <WO 9611947A1 IA>

may be used to detect P36 subsequent to initiation of dimerization by ΔP37.

A gp37-36 chimer was prepared similarly to the procedures described above, except that the restriction enzyme TaqI was used instead of HaeIII. Briefly, the 5' fragment resulting from TaqI digestion of gene 37 was ligated to the 3' fragment resulting from TaqI digestion of gene 36. This produced a construct encoding a gp37-36 chimer in which amino acids 1-48 of gp37 were fused to amino acids 100-221 of gp36. This construct was expressed in *E. coli* BL21(DE3), and the chimer was detected as an 18 kD protein. This gp37-36 chimer was found to inhibit the growth of wild type T4 when expression of the gp37-36 chimer was induced prior to infection (in an *in vitro* phage inhibition assay).

EXAMPLE 3

MUTATION OF THE GP37-36 CHIMER TO PRODUCE COMPLEMENTARY SUPPRESSORS

The goal of this construction is to produce two variants of a dimerizable P37-36 chimer: One in which the N-terminus of the polypeptide is mutated (A, designated *P37-36) and one in which the C-terminus of the polypeptide is mutated (B, designated P37-36*). The requirement is that the mutated *P37 N-terminus cannot form a joint with the wild-type P36 C- terminus, but only with the mutated *P36 N-terminus. The rationale is that A and B each cannot polymerize independently (as the parent P37-36 protein can), but can only associate with each other sequentially (i.e., P37-36* + *P37-36 --> P37-36*--*P37-36).

A second construct, *p37-P36*, is formed by recombining *P37-36 and P37-36* *in vitro*. When the monomers *gp37-36* and gp37-36 are mixed in the presence of P37 initiator, gp37-36 would dimerize and polymerize to (P37-36)n; similarly, *P37 would only catalyze the polymerization of *gp37-36* to (*P37-36*)n. In this case, the two chimers could be of different size and different primary sequence with different potential side-group

SDOCID: <WO___9611947A1_IA>

interactions, and could initiate attachment at different surfaces depending on the attachment specificity of P37.

The starting bacterial strain is a su° strain of *E. coli* (which lacks the ability to suppress amber mutations). When this strain is infected with a mutant T4 bacteriophage containing amber mutations in genes 35, 36, and 37, phage replication is incomplete, since the tail fiber proteins cannot be synthesized. When this strain is first transformed with a plasmid that directs the expression of the wild type gp35, gp36and gp37 genes and induced with IPTG, and subsequently infected with mutant phage, infectious phage particles are produced; this is evidenced by the appearance of "nibbled" colonies. Nibbled colonies do not appear round, with smooth edges, but rather have sectors missing. This is caused by attack of a microcolony by a single phage, which replicates and prevents the growth of the bacteria in the missing sector.

For the purposes of this construction, the 3'-terminal region of gene 36 (corresponding to the C-terminal region of gp36) is mutagenized with randomly doped oligonucleotides. Randomly doped oligonucleotides are prepared during chemical synthesis of oligonucleotides, by adding a trace amount (up to a few percent) of the other three nucleotides at a given position, so that the resulting oligonucleotide mix has a small percentage of incorrect nucleotides at that position. Incorporation of such oligonucleotides into the plasmid will result in random mutations (Hutchison et al., Methods.Enzymol. 202:356, 1991).

The mutagenized population of plasmids (containing, however, unmodified genes 36 and 37), is then transformed into the su° bacteria, followed by infection with the mutant T4 phage as above. In this case, the appearance of non-"nibbled" colonies indicates that the mutated gp36 C-termini can no longer interact with wild type P37 to form functional tail fibers. The putative gp36 phenotypes found in such non-nibbled colonies are checked for lack of dimeric N-termini by appropriate immunospecificity as outlined above,

SDOCID: <WO 9611947A1 IA>

and positive colonies are used as source of plasmid for the next step.

Several of these mutated plasmids are recovered and subjected to a second round of mutagenesis, this time using doped oligonucleotides that introduce random mutations into the N-terminal region of gp37 present on the same plasmid. Again, the (now doubly) mutagenized plasmids are transformed into the supo strain of *E. coli* and transformants are infected with the mutant T4 phage. At this stage, bacterial plates are screened for the re-appearance of "nibbled" colonies. A nibbled colony at this stage indicates that the phage has replicated by virtue of suppression of the non-functional gp36* mutation(s) by the *P37 mutation. In other words, such colonies must contain novel *P37 polypeptides that have now acquired the ability to interact with the P36* proteins encoded on the same plasmid.

The *P37-36 and P37-36* paired suppressor chimers (A and B as above) are then constructed in the same manner as described in Example 2. In this case, however, *P37 is used in place of wild type P37 and P36* is used in place of wild type P36. A *P37-36* chimer can now be made by restriction of *P37-36 and P37-36* and religation in the recombined order. The *P37-36* can be mixed with the P37-36 chimer, and the polymerization of each can be accomplished independently in the presence of the other. This is useful when the rod-like central portion of these chimers have been modified in different ways.

EXAMPLE 4

DESIGN, CONSTRUCTION AND EXPRESSION OF A gp36-34 CHIMER

The starting plasmid for this construction is one in which the vector containing gene 57 and the gene encoding gp36 is cloned immediately upstream (*i.e.*, 5') of the gene encoding gp34. The plasmid is digested with NdeI, which cuts after bp 219 of gene 36 and after bp 2594 of gene 34, thereby deleting the final 148 C-terminal codons from the pg36 moiety and the first 865 N-terminal codons from the gp34 moiety.

BNSDOCID: <WO___9611947A1_IA>

The reacti n mixture is id ntical to that described in Example 1, except that a different plasmid DNA is used, and the enzyme us d is NdeI (NEB). Ligation using T4 DNA ligase, bacterial transformation, and restriction analysis are also performed as in Example 1. This results in a new hybrid gene encoding a protein of 497 amino acids (73 N-terminal amino acids of gp36 and 424 C-terminal amino acids of gp34, numbered 866-1289.)

As an alternative, the starting plasmid is cut with SphI at bp 648 in gene 34, and the Exo-Size Deletion Kit (NEB) is used to create deletions as described above.

The resulting construct is then expressed in *E. coli* BL21(DE3) as described in Example 1. Successful expression of the gp36-34 chimer is evidenced by the appearance of a protein product of about 55,000 daltons. Preferably, the amino termini of the polypeptide homodimer have the capability of interacting with the gp35 protein, and then the carboxy termini have the capability of interacting with other attached gp35 molecules. Successful formation of the dimer can be detected by reaction with anti-P36 antibodies or by attachment of gp35 or by the *in vitro* phage inhibition assay described in Example 2.

EXAMPLE 5

ISOLATION OF THERMOLABILE PROTEINS FOR SELF-ASSEMBLY

Thermolabile structures can be utilized in nanostructures for: a) initiation of chimer polymerization (*e.g.*, gp37-36) at low temperature and subsequent inactivation of and separation from the initiator at high temperature; b) initiation of angle formation between P36 and gp35 (*e.g.*, variants of gp35 that have thermolabile attachment sites for P36 N-termini or P34 C-termini, a variant P36 that forms a thermolabile attachment to gp35, and a variant P34 with a thermolabile C-terminal attachment sit .) Thermolability may be reversible, permitting reattachment of the appr priate termini when the lower temperature is rest r d, or it may be irreversible.

WSDOCID: <WO___9611947A1 IA>

To create a variant gp37 that permits heat induced separation of the P36 -- P37 junction, the 5' end of gp37 DNA is randomly mutagenized using doped oligonucleotids as described above. The mutagenized DNA fragment is then recombined into T4 phage by infection of the cell containing the mutagenized DNA by a T4 phage containing two amber mutations flanking the mutagenized region. Following a low-multiplicity infection, non-amber phage are selected at low temperature on *E. coli* Su° at 30°C. The progeny of these plaques are resuspended in buffered and challenged by heating at 60°C. At this temperature, wild-type tail fibers remain intact and functional, whereas the thermolabile versions release the terminal P37 units and thus render those phage non-infectious.

At this stage, wild type phage are removed by: 1) adsorbing the wild type phage to sensitive bacteria and sedimenting (or filtering out) the bacteria with the adsorbed wild type phage; or 2) reacting the lysate with anti-P37 antibody, followed by immobilized Protein A and removal of adsorbed wild type phage. Either method leaves the noninfectious mutant phage particles in the supernatant fluid or filtrate, from which they can be recovered. The non-infectious phage lacking terminal P37 moieties (and probably the rest of the tail fibers as well) are then urea treated with 6M urea, and mixed with bacterial spheroplasts to permit infection at low multiplicity whereupon they replicate at low temperature and release progeny. Alternatively, infectious phage are reconstituted by *in vitro* incubation of the mutant phage with wild type P37 at 30°C; this is followed by infection of intact bacterial cells using the standard protocol. The latter method of infection specifically selects mutant phage in which the thermolability of the P36-P37 junction is reversible.

Using either method, the phage populations are subjected to multiple rounds of selection as above, after which individual phage particles are isolated by plaque purification at 30°C. Finally, the putative mutants ar

BNSDOCID: <WO___9611947A1_IA>

evaluated individually for the foll wing characteristics: 1) loss of infectivity after incubation at high temperatures (40-60°C), as measur d by a decr ase in titer; 2) loss of P37 after incubation at high temperature, as measured by decreas in binding of P37-specific antibody to phage particles; and 3) morphological changes in the tail fibers after incubation at high temperatures, as assessed by electron microscopy.

After mutants are isolated and their phenotypes confirmed, the P37 gene is sequenced. If the mutations localize to particular regions or residues, those sequences are targeted for site-directed mutagenesis to optimize the desired characteristics.

Finally, the mutant gene 37 is cloned into expression plasmids and expressed individually in *E. coli* as in Example 1. The mutant P37 dimers are then purified from bacterial extracts and used in *in vitro* assembly reactions.

In a similar fashion, mutant gp35 polypeptides can be isolated that exhibit a thermolabile interaction with th N-terminus of P36 or the C-terminus of P34. For thermolabile interaction with P34, phage are incubated at high temperature, resulting in the loss of the entire distal half of the tail fiber (*i.e.*, gp35-P36-P37). The only difference in the experimental protocol is that, in this case, 1) random mutagenesis is performed over the entire gp35 gene; 2) wild-type phage (and distal half-fibers from thermolabile mutants) are separated from thermolabile mutant phage that have been inactivated at high temperature (but still have proximal half tail fibers attached) by precipitating both the distal half-fibers and the phage particles containing intact tail fibers with any of the anti-distal half tail-fiber antibodies followed by Staphylococcal A-protein beads; 3) the mutant phage remaining in the supernatant are reactivated by incubation at low temperature with bacterial extracts containing wild type intact distal half fibers; and 4) stocks of therm labile gene,35 mutants grown at 30°C can be tested for reversibl thermolability by inactivation at 60°C and r incubation at 30°C. Inactivation is p rformed on a

NSDOCID: <WO__9611947A1 IA>

c ncentrated suspension of phage, and reincubation at 30°C is performed ither before or after dilution. If phage are successfully reactivated b fore, but not after, dilution, this indicates that their gp35 is reversibly thermolabile.

To create a gene 36 mutation with a thermolabile gp35--P36 linkage, the C-terminus of gene 36 is mutagenized as described above, and the mutant selected for reversibility. An alternative is to mutagenize gp35 to create a gene 35 mutant in which the gp35-P36 linkage will dissociate at 60°C. In this case, incubation with anti-gp35 antibodies can be used to precipitate the phage without P36-P37 and thus to separate them from the wild-type phage and distal half-tail fibers (P36-P37), since the variant gp35 will remain attached to P34.

EXAMPLE 6

ASSEMBLY OF ONE-DIMENSIONAL RODS

A. Simple Assembly: The P37-36 chimer described in Example 2 is capable of self-assembly, but requires a P37 initiator to bind the first unit of the rod. Therefore, a P37 or a ΔP37 dimer is either attached to a solid matrix or is free in solution to serve as an initiator. If the initiator is, attached to a solid matrix, a thermolabile P37 dimer is preferably used. Addition of an extract containing gp37-36, or the purified gp37-36 chimer, results in the assembly of linear multimers of increasing length. In the matrix-bound case, the final rods are released by a brief incubation at high temperature (40-60°C, depending on the characteristics of the particular thermolabile P37 variant.)

The ratio of initiator to gp37-36 can be varied, and the size distribution of the rods is measured by any of the following methods: 1) Size exclusion chromatography; 2) Increase in the viscosity of the solution; and 3) Direct measurement by electron microscopy.

B. Staged assembly: The P37-36 variants *P37-36 and P37-36* described in Example 3 cannot self-polymerize.

NSDOCID: <WO___9611947A1_IA>

This allows th staged assembly f rods of defined length, according to the following pr t c l:

1. Attach initiat r P37 (preferably thermolabile) to a matrix.

2. Add excess *gp37-36 to attach and oligomerize as P37-36 homooligomers to the N-terminus of P37.

3. Wash out unreacted *gp37-36 and flood with gp37-36*.

4. Wash out unreacted gp37-36* and flood with excess *gp37-36.

5. Repeat steps 2-4, n-1 times.

6. Release assembly from matrix by brief incubation at high temperature as above.

The linear dimensions of the protein rods in the batch will depend upon the lengths of the unit heterochimers and the number of cycles (n) of addition. This method has the advantage of insuring absolute reproducibility of rod length and a homogenous, monodisperse size distribution from one preparation to another.

EXAMPLE 7

STAGED ASSEMBLY OF POLYGONS

The following assembly strategy utilizes gp35 as an angle joint to allow the formation of polygons. For the purpose of this example, the angle formed by gp35 is assumed to be 137°. The rod unit comprises the P36-34 chimer described in Example 4, which is incapable of self-polymerization. The P36-34 homodimer is made from a bacterial clone in which both gp36-34 and gp57 are expressed. The gp57 can chaperone the homodimerization of gp36-34 to P36-34.

1. Initiator: The incomplete distal half fiber P36-37 is attached to a solid matrix by the P37 C-terminus. Thermolabile gp35 as describ d in Example 5 is then added to form the intact initiator.

2. Excess P36-34 chimer is added to attach a single P36-34. Following binding to the matrix via gp35, the unbound chimer is washed ut.

3. Wild-typ (*i*. ., n n-therm labile) gp35 is then added in excess. After incubation, the unbound material is washed out.

4. Steps 2 and 3 are repeated 7-8 times.

5. The assembly is released from the matrix by brief incubation at high temperature.

The released polymeric rod, 8 units long, will form a regular 8-sided polygon, whose sides comprise the P36-34 dimer and whose joints comprise the wild-type gp35 monomer. However, there will be some multimers of these 8 units bound as helices. When a unit does not close, but instead adds another to its terminus, the unit cannot close further and the helix can build in either direction. The direction of the first overlap also determines the handedness of the helix. Ten (or seven)-unit rods may form helices mor frequently than polygons since their natural angles are 144° (or 128.6°). The likelihood of closure of a regular polygon depends not only on the average angle of gp35 but also on its flexibility, which can be further manipulated by genetic or environmental modification.

The type of polygon that is formed using this protocol depends upon the length of rod units and the angle formed by the angle joint. For example, alternating rod units of different sizes can be used in step 2. In addition, variant gp35 polypeptides that form angles different than the natural angle of 137° can be used, allowing the formation of different regular polygons. Furthermore, for a given polygon with an even number of sides and equal angles, the sides in either half can be of any size provided the two halves are symmetric.

INSDOCID: <WO___9611947A1_IA>

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Goldberg, Edward B.

(ii) TITLE OF INVENTION: MATERIALS FOR THE PRODUCTION OF NANOMETER STRUCTURES AND USE THEREOF

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Pennie and Edmonds
(B) STREET: 1155 Avenue of the Americas
(C) CITY: New York
(D) STATE: New York
(E) COUNTRY: US
(F) ZIP: 10036

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: To be Assigned
(B) FILING DATE: 13-OCT-1995
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Misrock, S. Leslie
(B) REGISTRATION NUMBER: 18,872
(C) REFERENCE/DOCKET NUMBER: 8471-0005-999

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (212) 790-9090
(B) TELEFAX: 212-869-8864
(C) TELEX: 66441 PENNIE

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8855 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Bacteriophage T4

(vii) IMMEDIATE SOURCE:
(B) CLONE: TAIL FIBER GENES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TAGGAGCCCG GGAGAATGGC CGAGATTAAA AGAGAATTCA GAGCAGAAGA TGGTCTGGAC 60

ENSDOCID: <WO 9611947A1 IA>

GCAGGTGGTG ATAAAATAAT CAACGTAGCT TTAGCTGATC GTACCGTAGG AACTGACGGT 120

GTTAACGTTG ATTACTTAAT TCAAGAAAAC ACAGTTCAAC AGTATGATCC AACTCGTGGA 180

TATTTAAAAG ATTTTGTAAT CATTTATGAT AACCGCTTTT GGGCTGCTAT AAATGATATT 240

CCAAAACCAG CAGGAGCTTT TAATAGCGGA CGCTGGAGAG CATTACGTAC CGATGCTAAC 300

TGGATTACGG TTTCATCTGG TTCATATCAA TTAAAATCTG GTGAAGCAAT TTCGGTTAAC 360

ACCGCAGCTG GAAATGACAT CACGTTTACT TTACCATCTT CTCCAATTGA TGGTGATACT 420

ATCGTTCTCC AAGATATTGG AGGAAAACCT GGAGTTAACC AAGTTTTAAT TGTAGCTCCA 480

GTACAAAGTA TTGTAAACTT TAGAGGTGAA CAGGTACGTT CAGTACTAAT GACTCATCCA 540

AAGTCACAGC TAGTTTTAAT TTTTAGTAAT CGTCTGTGGC AAATGTATGT TGCTGATTAT 600

AGTAGAGAAG CTATAGTTGT AACACCAGCG AATACTTATC AAGCGCAATC CAACGATTTT 660

ATCGTACGTA GATTTACTTC TGCTGCACCA ATTAATGTCA AACTTCCAAG ATTTGCTAAT 720

CATGGCGATA TTATTAATTT CGTCGATTTA GATAAACTAA ATCCGCTTTA TCATACAATT 780

GTTACTACAT ACGATGAAAC GACTTCAGTA CAAGAAGTTG GAACTCATTC CATTGAAGGC 840

CGTACATCGA TTGACGGTTT CTTGATGTTT GATGATAATG AGAAATTATG GAGACTGTTT 900

GACGGGGATA GTAAAGCGCG TTTACGTATC ATAACGACTA ATTCAAACAT TCGTCCAAAT 960

GAAGAAGTTA TGGTATTTGG TGCGAATAAC GGAACAACTC AAACAATTGA GCTTAAGCTT 1020

CCAACTAATA TTTCTGTTGG TGATACTGTT AAAATTTCCA TGAATTACAT GAGAAAAGGA 1080

CAAACAGTTA AAATCAAAGC TGCTGATGAA GATAAAATTG CTTCTTCAGT TCAATTGCTG 1140

CAATTCCCAA AACGCTCAGA ATATCCACCT GAAGCTGAAT GGGTTACAGT TCAAGAATTA 1200

GTTTTTAACG ATGAAACTAA TTATGTTCCA GTTTTGGAGC TTGCTTACAT AGAAGATTCT 1260

GATGGAAAAT ATTGGGTTGT ACAGCAAAAC GTTCCAACTG TAGAAAGAGT AGATTCTTTA 1320

AATGATTCTA CTAGAGCAAG ATTAGGCGTA ATTGCTTTAG CTACACAAGC TCAAGCTAAT 1380

GTCGATTTAG AAAATTCTCC ACAAAAAGAA TTAGCAATTA CTCCAGAAAC GTTAGCTAAT 1440

CGTACTGCTA CAGAAACTCG CAGAGGTATT GCAAGAATAG CAACTACTGC TCAAGTGAAT 1500

CAGAACACCA CATTCTCTTT TGCTGATGAT ATTATCATCA CTCCTAAAAA GCTGAATGAA 1560

AGAACTGCTA CAGAAACTCG TAGAGGTGTC GCAGAAATTG CTACGCAGCA AGAAACTAAT 1620

GCAGGAACCG ATGATACTAC AATCATCACT CCTAAAAAGC TTCAAGCTCG TCAAGGTTCT 1680

GAATCATTAT CTGGTATTGT AACCTTTGTA TCTACTGCAG GTGCTACTCC AGCTTCTAGC 1740

CGTGAATTAA ATGGTACGAA TGTTTATAAT AAAAACACTG ATAATTTAGT TGTTTCACCT 1800

AAAGCTTTGG ATCAGTATAA AGCTACTCCA ACACAGCAAG GTGCAGTAAT TTTAGCAGTT 1860

GAAAGTGAAG TAATTGCTGG ACAAAGTCAG CAAGGATGGG CAAATGCTGT TGTAACGCCA 1920

GAAACGTTAC ATAAAAAGAC ATCAACTGAT GGAAGAATTG GTTTAATTGA AATTGCTACG 1980

CAAAGTGAAG TTAATACAGG AACTGATTAT ACTCGTGCAG TCACTCCTAA AACTTTAAAT 2040

GACCGTAGAG CAACTGAAAG TTTAAGTGGT ATAGCTGAAA TTGCTACACA AGTTGAATTC 2100

SDOCID: <WO___9611947A1_IA>

GACGCAGGCG TCGACGATAC TCGTATCTCT ACACCATTAA AAATTAAAAC CAGATTTAAT 2160

AGTACTGATC GTACTTCTGT TGTTGCTCTA TCTGGATTAG TTGAATCAGG AACTCTCTGG 2220

GACCATTATA CACTTAATAT TCTTGAAGCA AATGAGACAC AACGTGGTAC ACTTCGTGTA 2280

GCTACGCAGG TCGAAGCTGC TGCGGGAACA TTAGATAATG TTTTAATAAC TCCTAAAAAG 2340

CTTTTAGGTA CTAAATCTAC TGAAGCGCAA GAGGGTGTTA TTAAAGTTGC AACTCAGTCT 2400

GAAACTGTGA CTGGAACGTC AGCAAATACT GCTGTATCTC CAAAAAATTT AAAATGGATT 2460

CGCAGAGTG AACCTACTTG GGCAGCTACT ACTGCAATAA GAGGTTTTGT TAAAACTTCA 2520

TCTGGTTCAA TTACATTCGT TGGTAATGAT ACAGTCGGTT CTACCCAAGA TTTAGAACTG 2580

TATGAGAAAA ATAGCTATGC GGTATCACCA TATGAATTAA ACCGTGTATT AGCAAATTAT 2640

TTGCCACTAA AAGCAAAAGC TGCTGATACA AATTTATTGG ATGGTCTAGA TTCATCTCAG 2700

TTCATTCGTA GGGATATTGC ACAGACGGTT AATGGTTCAC TAACCTTAAC CCAACAAACG 2760

AATCTGAGTG CCCCTCTTGT ATCATCTAGT ACTGGTGAAT TTGGTGGTTC ATTGGCCGCT 2820

AATAGAACAT TTACCATCCG TAATACAGGA GCCCCGACTA GTATCGTTTT CGAAAAAGGT 2880

CCTGCATCCG GGGCAAATCC TGCACAGTCA ATGAGTATTC GTGTATGGGG TAACCAATTT 2940

GGCGGCGGTA GTGATACGAC CCGTTCGACA GTGTTTGAAG TTGGCGATGA CACATCTCAT 3000

CACTTTTATT CTCAACGTAA TAAAGACGGT AATATAGCGT TTAACATTAA TGGTACTGTA 3060

ATGCCAATAA ACATTAATGC TTCCGGTTTG ATGAATGTGA ATGGCACTGC AACATTCGGT 3120

CGTTCAGTTA CAGCCAATGG TGAATTCATC AGCAAGTCTG CAAATGCTTT TAGAGCAATA 3180

AACGGTGATT ACGGATTCTT TATTCGTAAT GATGCCTCTA ATACCTATTT TTTGCTCACT 3240

GCAGCCGGTG ATCAGACTGG TGGTTTTAAT GGATTACGCC CATTATTAAT TAATAATCAA 3300

TCCGGTCAGA TTACAATTGG TGAAGGCTTA ATCATTGCCA AAGGTGTTAC TATAAATTCA 3360

GGCGGTTTAA CTGTTAACTC GAGAATTCGT TCTCAGGGTA CTAAAACATC TGATTTATAT 3420

ACCCGTGCGC CAACATCTGA TACTGTAGGA TTCTGGTCAA TCGATATTAA TGATTCAGCC 3480

ACTTATAACC AGTTCCCGGG TTATTTTAAA ATGGTTGAAA AAACTAATGA AGTGACTGGG 3540

CTTCCATACT TAGAACGTGG CGAAGAAGTT AAATCTCCTG GTACACTGAC TCAGTTTGGT 3600

AACACACTTG ATTCGCTTTA CCAAGATTGG ATTACTTATC CAACGACGCC AGAAGCGCGT 3660

ACCACTCGCT GGACACGTAC ATGGCAGAAA ACCAAAAACT CTTGGTCAAG TTTTGTTCAG 3720

GTATTTGACG GAGGTAACCC TCCTCAACCA TCTGATATCG GTGCTTTACC ATCTGATAAT 3780

GCTACAATGG GGAATCTTAC TATTCGTGAT TTCTTGCGAA TTGGTAATGT TCGCATTGTT 3840

CCTGACCCAG TGAATAAAAC GGTTAAATTT GAATGGGTTG AATAAGAGGT ATTATGGAAA 3900

AATTTATGGC CGAGATTTGG ACAAGGATAT GTCCAAACGC CATTTTATCG GAAAGTAATT 3960

CAGTAAGATA TAAAATAAGT ATAGCGGGTT CTTGCCCGCT TTCTACAGCA GGACCATCAT 4020

ATGTTAAATT TCAGGATAAT CCTGTAGGAA GTCAAACATT TAGGCGCAGG CCTTCATTTA 4080

AGAGTTTTTG ACCCTTCCAC CGGAGCATTA GTTGATAGTA AGTCATATGC TTTTTCGACT 4140

TCAAATGATA CTACATCAGC TGCTTTTGTT AGTTTTCATG AATTCTTTGA CGAATAATCG 4200

AATTGTTGCT ATATTAACTA GTGGAAAGGT TAATTTTCCT CCTGAAGTAG TATCTTGGTT 4260

AAGAACCGCC GGAACGTCTG CCTTTCCATC TGATTCTATA TTGTCAAGAT TTGACGTATC 4320

ATATGCTGCT TTTTATACTT CTTCTAAAAG AGCTATCGCA TTAGAGCATG TTAAACTGAG 4380

TAATAGAAAA AGCACAGATG ATTATCAAAC TATTTTAGAT GTTGTATTTG ACAGTTTAGA 4440

AGATGTAGGA GCTACCGGGT TTCCAAGAAG AACGTATGAA AGTGTTGAGC AATTCATGTC 4500

GGCAGTTGGT GGAACTAATA ACGAAATTGC GAGATTGCCA ACTTCAGCTG CTATAAGTAA 4560

ATTATCTGAT TATAATTTAA TTCCTGGAGA TGTTCTTTAT CTTAAAGCTC AGTTATATGC 4620

TGATGCTGAT TTACTTGCTC TTGGAACTAC AAATATATCT ATCCGTTTTT ATAATGCATC 4680

TAACGGATAT ATTTCTTCAA CACAAGCTGA ATTTACTGGG CAAGCTGGGT CATGGGAATT 4740

AAAGGAAGAT TATGTAGTTG TTCCAGAAAA CGCAGTAGGA TTTACGATAT ACGCACAGAG 4800

AACTGCACAA GCTGGCCAAG GTGGCATGAG AAATTTAAGC TTTTCTGAAG TATCAAGAAA 4860

TGGCGGCATT TCGAAACCTG CTGAATTTGG CGTCAATGGT ATTCGTGTTA ATTATATCTG 4920

CGAATCCGCT TCACCTCCGG ATATAATGGT ACTTCCTACG CAAGCATCGT CTAAAACTGG 4980

TAAAGTGTTT GGGCAAGAAT TTAGAGAAGT TTAAATTGAG GGACCCTTCG GGTTCCCTTT 5040

TTCTTTATAA ATACTATTCA AATAAAGGGG CATACAATGG CTGATTTAAA AGTAGGTTCA 5100

ACAACTGGAG GCTCTGTCAT TTGGCATCAA GGAAATTTTC CATTGAATCC AGCCGGTGAC 5160

GATGTACTCT ATAAATCATT TAAAATATAT TCAGAATATA ACAAACCACA AGCTGCTGAT 5220

AACGATTTCG TTTCTAAAGC TAATGGTGGT ACTTATGCAT CAAAGGTAAC ATTTAACGCT 5280

GGCATTCAAG TCCCATATGC TCCAAACATC ATGAGCCCAT GCGGGATTTA TGGGGGTAAC 5340

GGTGATGGTG CTACTTTTGA TAAAGCAAAT ATCGATATTG TTTCATGGTA TGGCGTAGGA 5400

TTTAAATCGT CATTTGGTTC AACAGGCCGA ACTGTTGTAA TTAATACACG CAATGGTGAT 5460

ATTAACACAA AAGGTGTTGT GTCGGCAGCT GGTCAAGTAA GAAGTGGTGC GGCTGCTCCT 5520

ATAGCAGCGA ATGACCTTAC TAGAAAGGAC TATGTTGATG GAGCAATAAA TACTGTTACT 5580

GCAAATGCAA ACTCTAGGGT GCTACGGTCT GGTGACACCA TGACAGGTAA TTTAACAGCG 5640

CCAAACTTTT TCTCGCAGAA TCCTGCATCT CAACCCTCAC ACGTTCCACG ATTTGACCAA 5700

ATCGTAATTA AGGATTCTGT TCAAGATTTC GGCTATTATT AAGAGGACTT ATGGCTACTT 5760

TAAAACAAAT ACAATTTAAA AGAAGCAAAA TCGCAGGAAC ACGTCCTGCT GCTTCAGTAT 5820

TAGCCGAAGG TGAATTGGCT ATAAACTTAA AAGATAGAAC AATTTTTACT AAAGATGATT 5880

CAGGAAATAT CATCGATCTA GGTTTTGCTA AAGGCGGGCA AGTTGATGGC AACGTTACTA 5940

TTAACGGACT TTTGAGATTA AATGGCGATT ATGTACAAAC AGGTGGAATG ACTGTAAACG 6000

GACCCATTGG TTCTACTGAT GGCGTCACTG GAAAAATTTT CAGATCTACA CAGGGTTCAT 6060

TTTATGCAAG AGCAACAAAC GATACTTCAA ATGCCCATTT ATGGTTTGAA AATGCCGATG 6120

GCACTGAACG TGGCGTTATA TATGCTCGCC CTCAAACTAC AACTGACGGT GAAATACGCC 6180

BNSDOCID: <WO___9611947A1_IA>

TTAGGGTTAG ACAAGGAACA GGAAGCACTG CCAACAGTGA ATTCTATTTC CGCTCTATAA 6240

ATGGAGGCGA ATTTCAGGCT AACCGTATTT TAGCATCAGA TTCGTTAGTA ACAAAACGCA 6300

TTGCGGTTGA TACCGTTATT CATGATGCCA AAGCATTTGG ACAATATGAT TCTCACTCTT 6360

TGGTTAATTA TGTTTATCCT GGAACCGGTG AAACAAATGG TGTAAACTAT CTTCGTAAAG 6420

TTCGCGCTAA GTCCGGTGGT ACAATTTATC ATGAAATTGT TACTGCACAA ACAGGCCTGG 6480

CTGATGAAGT TTCTTGGTGG TCTGGTGATA CACCAGTATT TAAACTATAC GGTATTCGTG 6540

ACGATGGCAG AATGATTATC CGTAATAGCC TTGCATTAGG TACATTCACT ACAAATTTCC 6600

CGTCTAGTGA TTATGGCAAC GTCGGTGTAA TGGGCGATAA GTATCTTGTT CTCGGCGACA 6660

CTGTAACTGG CTTGTCATAC AAAAAAACTG GTGTATTTGA TCTAGTTGGC GGTGGATATT 6720

CTGTTGCTTC TATTACTCCT GACAGTTTCC GTAGTACTCG TAAAGGTATA TTTGGTCGTT 6780

CTGAGGACCA AGGCGCAACT TGGATAATGC CTGGTACAAA TGCTGCTCTC TTGTCTGTTC 6840

AAACACAAGC TGATAATAAC AATGCTGGAG ACGGACAAAC CCATATCGGG TACAATGCTG 6900

GCGGTAAAAT GAACCACTAT TTCCGTGGTA CAGGTCAGAT GAATATCAAT ACCCAACAAG 6960

GTATGGAAAT TAACCCGGGT ATTTTGAAAT TGGTAACTGG CTCTAATAAT GTACAATTTT 7020

ACGCTGACGG AACTATTTCT TCCATTCAAC CTATTAAATT AGATAACGAG ATATTTTTAA 7080

CTAAATCTAA TAATACTGCG GGTCTTAAAT TTGGAGCTCC TAGCCAAGTT GATGGCACAA 7140

GGACTATCCA ATGGAACGGT GGTACTCGCG AAGGACAGAA TAAAAACTAT GTGATTATTA 7200

AAGCATGGGG TAACTCATTT AATGCCACTG GTGATAGATC TCGCGAAACG GTTTTCCAAG 7260

TATCAGATAG TCAAGGATAT TATTTTTATG CTCATCGTAA AGCTCCAACC GGCGACGAAA 7320

CTATTGGACG TATTGAAGCT CAATTTGCTG GGGATGTTTA TGCTAAAGGT ATTATTGCCA 7380

ACGGAAATTT TAGAGTTGTT GGGTCAAGCG CTTTAGCCGG CAATGTTACT ATGTCTAACG 7440

GTTTGTTTGT CCAAGGTGGT TCTTCTATTA CTGGACAAGT TAAAATTGGC GGAACAGCAA 7500

ACGCACTGAG AATTTGGAAC GCTGAATATG GTGCTATTTT CCGTCGTTCG GAAAGTAACT 7560

TTTATATTAT TCCAACCAAT CAAAATGAAG GAGAAAGTGG AGACATTCAC AGCTCTTTGA 7620

GACCTGTGAG AATAGGATTA AACGATGGCA TGGTTGGGTT AGGAAGAGAT TCTTTTATAG 7680

TAGATCAAAA TAATGCTTTA ACTACGATAA ACAGTAACTC TCGCATTAAT GCCAACTTTA 7740

GAATGCAATT GGGGCAGTCG GCATACATTG ATGCAGAATG TACTGATGCT GTTCGCCCGG 7800

CGGGTGCAGG TTCATTTGCT TCCCAGAATA ATGAAGACGT CCGTGCGCCG TTCTATATGA 7860

ATATTGATAG AACTGATGCT AGTGCATATG TTCCTATTTT GAAACAACGT TATGTTCAAG 7920

GCAATGGCTG CTATTCATTA GGGACTTTAA TTAATAATGG TAATTTCCGA GTTCATTACC 7980

ATGGCGGCGG AGATAACGGT TCTACAGGTC CACAGACTGC TGATTTTGGA TGGGAATTTA 8040

TTAAAAACGG TGATTTTATT TCACCTCGCG ATTTAATAGC AGGCAAAGTC AGATTTGATA 8100

GAACTGGTAA TATCACTGGT GGTTCTGGTA ATTTTGCTAA CTTAAACAGT ACAATTGAAT 8160

CACTTAAAAC TGATATCATG TCGAGTTACC CAATTGGTGC TCCGATTCCT TGGCCGAGT 8220

NSDOCID: <WO___9611947A1 IA>

ATTCAGTTCC TGCTGGATTT GCTTTGATGG AAGGTCAGAC CTTTGATAAG TCCGCATATC 8280

CAAAGTTAGC TGTTGCATAT CCTAGCGGTG TTATTCCAGA TATGCGCGGG CAAACTATCA 8340

AGGGTAAACC AAGTGGTCGT GCTGTTTTGA GCGCTGAGGC AGATGGTGTT AAGGCTCATA 8400

GCCATAGTGC ATCGGCTTCA AGTACTGACT TAGGTACTAA AACCACATCA AGCTTTGACT 8460

ATGGTACGAA GGGAACTAAC AGTACGGGTG GACACACTCA CTCTGGTAGT GGTTCTACTA 8520

GCACAAATGG TGAGCACAGC CACTACATCG AGGCATGGAA TGGTACTGGT GTAGGTGGTA 8580

ATAAGATGTC ATCATATGCC ATATCATACA GGGCGGGTGG GAGTAACACT AATGCAGCAG 8640

GGAACCACAG TCACACTTTC TCTTTTGGGA CTAGCAGTGC TGGCGACCAT TCCCACTCTG 8700

TAGGTATTGG TGCTCATACC CACACGGTAG CAATTGGATC ACATGGTCAT ACTATCACTG 8760

TAAATAGTAC AGGTAATACA GAAAACACGG TTAAAAACAT TGCTTTTAAC TATATCGTTC 8820

GTTTAGCATA AGGAGAGGGG CTTCGGCCCT TCTAA 8855

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1289 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Bacteriophage T4

(vii) IMMEDIATE SOURCE:
(B) CLONE: p34 amino acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Ala Glu Ile Lys Arg Glu Phe Arg Ala Glu Asp Gly Leu Asp Ala
1 5 10 15

Gly Gly Asp Lys Ile Ile Asn Val Ala Leu Ala Asp Arg Thr Val Gly
20 25 30

Thr Asp Gly Val Asn Val Asp Tyr Leu Ile Gln Glu Asn Thr Val Gln
35 40 45

Gln Tyr Asp Pro Thr Arg Gly Tyr Leu Lys Asp Phe Val Ile Ile Tyr
50 55 60

Asp Asn Arg Phe Trp Ala Ala Ile Asn Asp Ile Pro Lys Pro Ala Gly
65 70 75 80

Ala Phe Asn Ser Gly Arg Trp Arg Ala Leu Arg Thr Asp Ala Asn Trp
85 90 95

Ile Thr Val Ser Ser Gly Ser Tyr Gln Leu Lys Ser Gly Glu Ala Ile
100 105 110

Ser Val Asn Thr Ala Ala Gly Asn Asp Ile Thr Phe Thr Leu Pro Ser
115 120 125

Ser Pro Ile Asp Gly Asp Thr Ile Val Leu Gln Asp Ile Gly Gly Lys
130 135 140

ISDOCID: <WO___9611947A1_IA>

Pro Gly Val Asn Gln Val L u Ile Val Ala Pro Val Gln Ser Ile Val
145 150 155 160

Asn Phe Arg Gly lu Gln Val Arg Ser Val L u Met Thr His Pro Lys
165 170 175

Ser Gln Leu Val Leu Ile Phe Ser Asn Arg Leu Trp Gln Met Tyr Val
180 185 190

Ala Asp Tyr Ser Arg Glu Ala Ile Val Val Thr Pro Ala Asn Thr Tyr
195 200 205

Gln Ala Gln Ser Asn Asp Phe Ile Val Arg Arg Phe Thr Ser Ala Ala
210 215 220

Pro Ile Asn Val Lys Leu Pro Arg Phe Ala Asn His Gly Asp Ile Ile
225 230 235 240

Asn Phe Val Asp Leu Asp Lys Leu Asn Pro Leu Tyr His Thr Ile Val
245 250 255

Thr Thr Tyr Asp Glu Thr Thr Ser Val Gln Glu Val Gly Thr His Ser
260 265 270

Ile Glu Gly Arg Thr Ser Ile Asp Gly Phe Leu Met Phe Asp Asp Asn
275 280 285

Glu Lys Leu Trp Arg Leu Phe Asp Gly Asp Ser Lys Ala Arg Leu Arg
290 295 300

Ile Ile Thr Thr Asn Ser Asn Ile Arg Pro Asn Glu Glu Val Met Val
305 310 315 320

Phe Gly Ala Asn Asn Gly Thr Thr Gln Thr Ile Glu Leu Lys Leu Pro
325 330 335

Thr Asn Ile Ser Val Gly Asp Thr Val Lys Ile Ser Met Asn Tyr Met
340 345 350

Arg Lys Gly Gln Thr Val Lys Ile Lys Ala Ala Asp Glu Asp Lys Ile
355 360 365

Ala Ser Ser Val Gln Leu Leu Gln Phe Pro Lys Arg Ser Glu Tyr Pro
370 375 380

Pro Glu Ala Glu Trp Val Thr Val Gln Glu Leu Val Phe Asn Asp Glu
385 390 395 400

Thr Asn Tyr Val Pro Val Leu Glu Leu Ala Tyr Ile Glu Asp Ser Asp
405 410 415

Gly Lys Tyr Trp Val Val Gln Gln Asn Val Pro Thr Val Glu Arg Val
420 425 430

Asp Ser Leu Asn Asp Ser Thr Arg Ala Arg Leu Gly Val Ile Ala Leu
435 440 445

Ala Thr Gln Ala Gln Ala Asn Val Asp Leu Glu Asn Ser Pro Gln Lys
450 455 460

Glu Leu Ala Ile Thr Pro Glu Thr Leu Ala Asn Arg Thr Ala Thr Glu
465 470 475 480

Thr Arg Arg Gly Ile Ala Arg Ile Ala Thr Thr Ala Gln Val Asn Gln
485 490 495

Asn Thr Thr Phe Ser Phe Ala Asp Asp Ile Il Il Thr Pro Lys Lys

3NSDOCID: <WO 9611947A1 IA>

500 505 510

Leu Asn Glu Arg Thr Ala Thr Glu Thr Arg Arg Gly Val Ala Glu Ile
515 520 525

Ala Thr Gln Gln Glu Thr Asn Ala Gly Thr Asp Asp Thr Thr Ile Il
530 535 540

Thr Pro Lys Lys Leu Gln Ala Arg Gln Gly Ser Glu Ser Leu Ser Gly
545 550 555 560

Ile Val Thr Phe Val Ser Thr Ala Gly Ala Thr Pro Ala Ser Ser Arg
565 570 575

Glu Leu Asn Gly Thr Asn Val Tyr Asn Lys Asn Thr Asp Asn Leu Val
580 585 590

Val Ser Pro Lys Ala Leu Asp Gln Tyr Lys Ala Thr Pro Thr Gln Gln
595 600 605

Gly Ala Val Ile Leu Ala Val Glu Ser Glu Val Ile Ala Gly Gln Ser
610 615 620

Gln Gln Gly Trp Ala Asn Ala Val Val Thr Pro Glu Thr Leu His Lys
625 630 635 640

Lys Thr Ser Thr Asp Gly Arg Ile Gly Leu Ile Glu Ile Ala Thr Gln
645 650 655

Ser Glu Val Asn Thr Gly Thr Asp Tyr Thr Arg Ala Val Thr Pro Lys
660 665 670

Thr Leu Asn Asp Arg Arg Ala Thr Glu Ser Leu Ser Gly Ile Ala Glu
675 680 685

Ile Ala Thr Gln Val Glu Phe Asp Ala Gly Val Asp Asp Thr Arg Ile
690 695 700

Ser Thr Pro Leu Lys Ile Lys Thr Arg Phe Asn Ser Thr Asp Arg Thr
705 710 715 720

Ser Val Val Ala Leu Ser Gly Leu Val Glu Ser Gly Thr Leu Trp Asp
725 730 735

His Tyr Thr Leu Asn Ile Leu Glu Ala Asn Glu Thr Gln Arg Gly Thr
740 745 750

Leu Arg Val Ala Thr Gln Val Glu Ala Ala Ala Gly Thr Leu Asp Asn
755 760 765

Val Leu Ile Thr Pro Lys Lys Leu Leu Gly Thr Lys Ser Thr Glu Ala
770 775 780

Gln Glu Gly Val Ile Lys Val Ala Thr Gln Ser Glu Thr Val Thr Gly
785 790 795 800

Thr Ser Ala Asn Thr Ala Val Ser Pro Lys Asn Leu Lys Trp Ile Ala
805 810 815

Gln Ser Glu Pro Thr Trp Ala Ala Thr Thr Ala Ile Arg Gly Phe Val
820 825 830

Lys Thr Ser Ser Gly Ser Ile Thr Phe Val Gly Asn Asp Thr Val Gly
835 840 845

Ser Thr Gln Asp Leu Glu Leu Tyr Glu Lys Asn Ser Tyr Ala Val Ser
850 855 860

BNSDOCID: <WO___9611947A1_IA>

Pro Tyr Glu Leu Asn Arg Val Leu Ala Asn Tyr Leu Pro Leu Lys Ala
865 870 875 880

Lys Ala Ala Asp Thr Asn Leu Leu Asp Gly Leu Asp Ser Ser Gln Phe
885 890 895

Ile Arg Arg Asp Ile Ala ln Thr Val Asn Gly Ser Leu Thr Leu Thr
900 905 910

Gln Gln Thr Asn Leu Ser Ala Pro Leu Val Ser Ser Ser Thr Gly Glu
915 920 925

Phe Gly Gly Ser Leu Ala Ala Asn Arg Thr Phe Thr Ile Arg Asn Thr
930 935 940

Gly Ala Pro Thr Ser Ile Val Phe Glu Lys Gly Pro Ala Ser Gly Ala
945 950 955 960

Asn Pro Ala Gln Ser Met Ser Ile Arg Val Trp Gly Asn Gln Phe Gly
965 970 975

Gly Gly Ser Asp Thr Thr Arg Ser Thr Val Phe Glu Val Gly Asp Asp
980 985 990

Thr Ser His His Phe Tyr Ser Gln Arg Asn Lys Asp Gly Asn Ile Ala
995 1000 1005

Phe Asn Ile Asn Gly Thr Val Met Pro Ile Asn Ile Asn Ala Ser Gly
1010 1015 1020

Leu Met Asn Val Asn Gly Thr Ala Thr Phe Gly Arg Ser Val Thr Ala
1025 1030 1035 1040

Asn Gly Glu Phe Ile Ser Lys Ser Ala Asn Ala Phe Arg Ala Ile Asn
1045 1050 1055

Gly Asp Tyr Gly Phe Phe Ile Arg Asn Asp Ala Ser Asn Thr Tyr Phe
1060 1065 1070

Leu Leu Thr Ala Ala Gly Asp Gln Thr Gly Gly Phe Asn Gly Leu Arg
1075 1080 1085

Pro Leu Leu Ile Asn Asn Gln Ser Gly Gln Ile Thr Ile Gly Glu Gly
1090 1095 1100

Leu Ile Ile Ala Lys Gly Val Thr Ile Asn Ser Gly Gly Leu Thr Val
1105 1110 1115 1120

Asn Ser Arg Ile Arg Ser Gln Gly Thr Lys Thr Ser Asp Leu Tyr Thr
1125 1130 1135

Arg Ala Pro Thr Ser Asp Thr Val Gly Phe Trp Ser Ile Asp Ile Asn
1140 1145 1150

Asp Ser Ala Thr Tyr Asn Gln Phe Pro Gly Tyr Phe Lys Met Val Glu
1155 1160 1165

Lys Thr Asn Glu Val Thr Gly Leu Pro Tyr Leu Glu Arg Gly Glu Glu
1170 1175 1180

Val Lys Ser Pro Gly Thr Leu Thr Gln Phe Gly Asn Thr Leu Asp Ser
1185 1190 1195 1200

Leu Tyr Gln Asp Trp Ile Thr Tyr Pr Thr Thr Pro Glu Ala Arg Thr
1205 1210 1215

Thr Arg Trp Thr Arg Thr Trp Gln Lys Thr Lys Asn Ser Trp S r Ser

NSDOCID: <WO 9611947A1 IA-

1220 1225 1230

Phe Val Gln Val Phe Asp Gly Gly Asn Pro Pro Gln Pro Ser Asp Ile
1235 1240 1245

Gly Ala L u Pro S r Asp Asn Ala Thr Met Gly Asn Leu Thr Ile Arg
1250 1255 1260

Asp Phe Leu Arg Ile Gly Asn Val Arg Ile Val Pro Asp Pro Val Asn
1265 1270 1275 1280

Lys Thr Val Lys Phe Glu Trp Val Glu
1285

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Bacteriophage T4

(vii) IMMEDIATE SOURCE:
(B) CLONE: ORF X amino acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Glu Lys Phe Met Ala Glu Ile Trp Thr Arg Ile Cys Pro Asn Ala
1 5 10 15

Ile Leu Ser Glu Ser Asn Ser Val Arg Tyr Lys Ile Ser Ile Ala Gly
20 25 30

Ser Cys Pro Leu Ser Thr Ala Gly Pro Ser Tyr Val Lys Phe Gln Asp
35 40 45

Asn Pro Val Gly Ser Gln Thr Phe Arg Arg Arg Pro Ser Phe Lys Ser
50 55 60

Phe
65

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 295 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Bacteriophage T4

(vii) IMMEDIATE SOURCE:
(B) CLONE: p35 amino acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Leu Phe Arg Leu Gln Met Ile Leu His Gln Leu Leu Leu Leu Val
1 5 10 15

NSDOCID: <WO___9611947A1_IA>

Phe Met Asn Ser Leu Thr Asn Asn Arg Ile Val Ala Ile Leu Thr Ser
20 25 30

Gly Lys Val Asn Ph Pro Pro Glu Val Val Ser Trp Leu Arg Thr Ala
35 40 45

Gly Thr S r Ala Phe Pro Ser Asp Ser Ile Leu Ser Arg Phe Asp Val
50 55 60

Ser Tyr Ala Ala Phe Tyr Thr Ser Ser Lys Arg Ala Ile Ala Leu Glu
65 70 75 80

His Val Lys Leu Ser Asn Arg Lys Ser Thr Asp Asp Tyr Gln Thr Ile
85 90 95

Leu Asp Val Val Phe Asp Ser Leu Glu Asp Val Gly Ala Thr Gly Phe
100 105 110

Pro Arg Arg Thr Tyr Glu Ser Val Glu Gln Phe Met Ser Ala Val Gly
115 120 125

Gly Thr Asn Asn Glu Ile Ala Arg Leu Pro Thr Ser Ala Ala Ile Ser
130 135 140

Lys Leu Ser Asp Tyr Asn Leu Ile Pro Gly Asp Val Leu Tyr Leu Lys
145 150 155 160

Ala Gln Leu Tyr Ala Asp Ala Asp Leu Leu Ala Leu Gly Thr Thr Asn
165 170 175

Ile Ser Ile Arg Phe Tyr Asn Ala Ser Asn Gly Tyr Ile Ser Ser Thr
180 185 190

Gln Ala Glu Phe Thr Gly Gln Ala Gly Ser Trp Glu Leu Lys Glu Asp
195 200 205

Tyr Val Val Val Pro Glu Asn Ala Val Gly Phe Thr Ile Tyr Ala Gln
210 215 220

Arg Thr Ala Gln Ala Gly Gln Gly Gly Met Arg Asn Leu Ser Phe Ser
225 230 235 240

Glu Val Ser Arg Asn Gly Gly Ile Ser Lys Pro Ala Glu Phe Gly Val
245 250 255

Asn Gly Ile Arg Val Asn Tyr Ile Cys Glu Ser Ala Ser Pro Pro Asp
260 265 270

Ile Met Val Leu Pro Thr Gln Ala Ser Ser Lys Thr Gly Lys Val Phe
275 280 285

Gly Gln Glu Phe Arg Glu Val
290 295

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Bacteriophag T4

(vii) IMMEDIATE SOURCE:

SDOCID: <WO 9611947A1 IA>

(B) CLONE: p36 amino acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
Met Ala Asp Leu Lys Val Gly Ser Thr Thr Gly Gly Ser Val Ile Trp
1               5                  10                  15

His Gln Gly Asn Phe Pro Leu Asn Pro Ala Gly Asp Asp Val Leu Tyr
            20                  25                  30

Lys Ser Phe Lys Ile Tyr Ser Glu Tyr Asn Lys Pro Gln Ala Ala Asp
        35                  40                  45

Asn Asp Phe Val Ser Lys Ala Asn Gly Gly Thr Tyr Ala Ser Lys Val
    50                  55                  60

Thr Phe Asn Ala Gly Ile Gln Val Pro Tyr Ala Pro Asn Ile Met Ser
65                  70                  75                  80

Pro Cys Gly Ile Tyr Gly Gly Asn Gly Asp Gly Ala Thr Phe Asp Lys
                85                  90                  95

Ala Asn Ile Asp Ile Val Ser Trp Tyr Gly Val Gly Phe Lys Ser Ser
            100                 105                 110

Phe Gly Ser Thr Gly Arg Thr Val Val Ile Asn Thr Arg Asn Gly Asp
        115                 120                 125

Ile Asn Thr Lys Gly Val Val Ser Ala Ala Gly Gln Val Arg Ser Gly
    130                 135                 140

Ala Ala Ala Pro Ile Ala Ala Asn Asp Leu Thr Arg Lys Asp Tyr Val
145                 150                 155                 160

Asp Gly Ala Ile Asn Thr Val Thr Ala Asn Ala Asn Ser Arg Val Leu
                165                 170                 175

Arg Ser Gly Asp Thr Met Thr Gly Asn Leu Thr Ala Pro Asn Phe Phe
            180                 185                 190

Ser Gln Asn Pro Ala Ser Gln Pro Ser His Val Pro Arg Phe Asp Gln
        195                 200                 205

Ile Val Ile Lys Asp Ser Val Gln Asp Phe Gly Tyr Tyr
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1026 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Bacteriophage T4

(vii) IMMEDIATE SOURCE:
(B) CLONE: p37 amino acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Ala Thr L u Lys Gln Il  Gln Phe Lys Arg Ser Lys Ile Ala Gly
1               5                  10                  15
```


NSDOCID: <WO__9611947A1_IA>

Thr Arg Pr Ala Ala Ser Val Leu Ala Glu Gly Glu Leu Ala Ile Asn
20 25 30

Leu Lys Asp Arg Thr Il Phe Thr Lys Asp Asp Ser Gly Asn Ile Ile
35 40 45

Asp Leu Gly Phe Ala Lys Gly Gly Gln Val Asp Gly Asn Val Thr Ile
50 55 60

Asn Gly Leu Leu Arg Leu Asn Gly Asp Tyr Val Gln Thr Gly Gly Met
65 70 75 80

Thr Val Asn Gly Pro Ile Gly Ser Thr Asp Gly Val Thr Gly Lys Ile
85 90 95

Phe Arg Ser Thr Gln Gly Ser Phe Tyr Ala Arg Ala Thr Asn Asp Thr
100 105 110

Ser Asn Ala His Leu Trp Phe Glu Asn Ala Asp Gly Thr Glu Arg Gly
115 120 125

Val Ile Tyr Ala Arg Pro Gln Thr Thr Thr Asp Gly Glu Ile Arg Leu
130 135 140

Arg Val Arg Gln Gly Thr Gly Ser Thr Ala Asn Ser Glu Phe Tyr Phe
145 150 155 160

Arg Ser Ile Asn Gly Gly Glu Phe Gln Ala Asn Arg Ile Leu Ala Ser
165 170 175

Asp Ser Leu Val Thr Lys Arg Ile Ala Val Asp Thr Val Ile His Asp
180 185 190

Ala Lys Ala Phe Gly Gln Tyr Asp Ser His Ser Leu Val Asn Tyr Val
195 200 205

Tyr Pro Gly Thr Gly Glu Thr Asn Gly Val Asn Tyr Leu Arg Lys Val
210 215 220

Arg Ala Lys Ser Gly Gly Thr Ile Tyr His Glu Ile Val Thr Ala Gln
225 230 235 240

Thr Gly Leu Ala Asp Glu Val Ser Trp Trp Ser Gly Asp Thr Pro Val
245 250 255

Phe Lys Leu Tyr Gly Ile Arg Asp Asp Gly Arg Met Ile Ile Arg Asn
260 265 270

Ser Leu Ala Leu Gly Thr Phe Thr Thr Asn Phe Pro Ser Ser Asp Tyr
275 280 285

Gly Asn Val Gly Val Met Gly Asp Lys Tyr Leu Val Leu Gly Asp Thr
290 295 300

Val Thr Gly Leu Ser Tyr Lys Lys Thr Gly Val Phe Asp Leu Val Gly
305 310 315 320

Gly Gly Tyr Ser Val Ala Ser Ile Thr Pro Asp Ser Phe Arg Ser Thr
325 330 335

Arg Lys Gly Ile Phe Gly Arg Ser Glu Asp Gln Gly Ala Thr Trp Ile
340 345 350

Met Pro Gly Thr Asn Ala Ala Leu Leu Ser Val Gln Thr Gln Ala Asp
355 360 365

Asn Asn Asn Ala Gly Asp Gly Gln Thr His Il Gly Tyr Asn Ala Gly

SUSDOCID <WO___9611947A1 IA>

370 375 380

Gly Lys Met Asn His Tyr Phe Arg Gly Thr Gly Gln Met Asn Ile Asn
385 390 395 400

Thr Gln Gln Gly Met Glu Il Asn Pro Gly Ile Leu Lys Leu Val Thr
405 410 415

Gly Ser Asn Asn Val Gln Ph Tyr Ala Asp Gly Thr Ile Ser Ser Ile
420 425 430

Gln Pro Ile Lys Leu Asp Asn Glu Ile Phe Leu Thr Lys Ser Asn Asn
435 440 445

Thr Ala Gly Leu Lys Phe Gly Ala Pro Ser Gln Val Asp Gly Thr Arg
450 455 460

Thr Ile Gln Trp Asn Gly Gly Thr Arg Glu Gly Gln Asn Lys Asn Tyr
465 470 475 480

Val Ile Ile Lys Ala Trp Gly Asn Ser Phe Asn Ala Thr Gly Asp Arg
485 490 495

Ser Arg Glu Thr Val Phe Gln Val Ser Asp Ser Gln Gly Tyr Tyr Phe
500 505 510

Tyr Ala His Arg Lys Ala Pro Thr Gly Asp Glu Thr Ile Gly Arg Ile
515 520 525

Glu Ala Gln Phe Ala Gly Asp Val Tyr Ala Lys Gly Ile Ile Ala Asn
530 535 540

Gly Asn Phe Arg Val Val Gly Ser Ser Ala Leu Ala Gly Asn Val Thr
545 550 555 560

Met Ser Asn Gly Leu Phe Val Gln Gly Gly Ser Ser Ile Thr Gly Gln
565 570 575

Val Lys Ile Gly Gly Thr Ala Asn Ala Leu Arg Ile Trp Asn Ala Glu
580 585 590

Tyr Gly Ala Ile Phe Arg Arg Ser Glu Ser Asn Phe Tyr Ile Ile Pro
595 600 605

Thr Asn Gln Asn Glu Gly Glu Ser Gly Asp Ile His Ser Ser Leu Arg
610 615 620

Pro Val Arg Ile Gly Leu Asn Asp Gly Met Val Gly Leu Gly Arg Asp
625 630 635 640

Ser Phe Ile Val Asp Gln Asn Asn Ala Leu Thr Thr Ile Asn Ser Asn
645 650 655

Ser Arg Ile Asn Ala Asn Phe Arg Met Gln Leu Gly Gln Ser Ala Tyr
660 665 670

Ile Asp Ala Glu Cys Thr Asp Ala Val Arg Pro Ala Gly Ala Gly Ser
675 680 685

Phe Ala Ser Gln Asn Asn Glu Asp Val Arg Ala Pro Phe Tyr Met Asn
690 695 700

Ile Asp Arg Thr Asp Ala Ser Ala Tyr Val Pro Ile Leu Lys Gln Arg
705 710 715 720

Tyr Val Gln Gly Asn Gly Cys Tyr Ser Leu Gly Thr Leu Ile Asn Asn
725 730 735

3NSDOCID: <WO___9611947A1_IA>

Gly Asn Phe Arg Val His Tyr His Gly Gly Gly Asp Asn Gly Ser Thr
740 745 750

Gly Pr Gln Thr Ala Asp Ph Gly Trp Glu Phe Ile Lys Asn Gly Asp
755 760 765

Phe Ile Ser Pr Arg Asp Leu Il Ala Gly Lys Val Arg Phe Asp Arg
770 775 780

Thr Gly Asn Ile Thr Gly Gly Ser Gly Asn Phe Ala Asn Leu Asn Ser
785 790 795 800

Thr Ile Glu Ser Leu Lys Thr Asp Ile Met Ser Ser Tyr Pro Ile Gly
805 810 815

Ala Pro Ile Pro Trp Pro Ser Asp Ser Val Pro Ala Gly Phe Ala Leu
820 825 830

Met Glu Gly Gln Thr Phe Asp Lys Ser Ala Tyr Pro Lys Leu Ala Val
835 840 845

Ala Tyr Pro Ser Gly Val Ile Pro Asp Met Arg Gly Gln Thr Ile Lys
850 855 860

Gly Lys Pro Ser Gly Arg Ala Val Leu Ser Ala Glu Ala Asp Gly Val
865 870 875 880

Lys Ala His Ser His Ser Ala Ser Ala Ser Ser Thr Asp Leu Gly Thr
885 890 895

Lys Thr Thr Ser Ser Phe Asp Tyr Gly Thr Lys Gly Thr Asn Ser Thr
900 905 910

Gly Gly His Thr His Ser Gly Ser Gly Ser Thr Ser Thr Asn Gly Glu
915 920 925

His Ser His Tyr Ile Glu Ala Trp Asn Gly Thr Gly Val Gly Gly Asn
930 935 940

Lys Met Ser Ser Tyr Ala Ile Ser Tyr Arg Ala Gly Gly Ser Asn Thr
945 950 955 960

Asn Ala Ala Gly Asn His Ser His Thr Phe Ser Phe Gly Thr Ser Ser
965 970 975

Ala Gly Asp His Ser His Ser Val Gly Ile Gly Ala His Thr His Thr
980 985 990

Val Ala Ile Gly Ser His Gly His Thr Ile Thr Val Asn Ser Thr Gly
995 1000 1005

Asn Thr Glu Asn Thr Val Lys Asn Ile Ala Phe Asn Tyr Ile Val Arg
1010 1015 1020

Leu Ala
1025

SDOCID: <WO___9611947A1 IA>

What is claimed is:

1. An isolated polypeptide consisting essentially of the gp37 tail fiber protein of bacteriophage T4 lacking amino acids 99-496 (SEQ ID NO:6) when numbered from the amino terminus, wherein said polypeptide has the capability to form dimers and interact with the P36 protein oligomer of bacteriophage T4.

2. An isolated polypeptide consisting essentially of a fusion protein between the gp36 and gp37 proteins of bacteriophage T4, wherein amino acid residues 1-242 of gp37 (SEQ ID NO:6) are fused in proper reading frame to amino acid residues 118-221 of gp36 (SEQ ID NO:5).

3. The polypeptide of claim 2 wherein a plurality of carboxy termini of said polypeptide have the capability of interacting with the amino terminus of the P37 protein oligomer of bacteriophage T4 and to form an attached oligomer and the amino termini of the oligomer of said polypeptide have the capability of interacting with the carboxy termini of gp36 polypeptides of bacteriophage T4.

4. An isolated polypeptide oligomer consisting essentially of two gp37 polypeptides of bacteriophage T4, wherein the amino termini of said oligomer lack the capability of interacting with the carboxy termini of gp36 polypeptides of bacteriophage T4.

5. An isolated polypeptide oligomer consisting essentially of the P37 protein of bacteriophage T4, wherein the amino termini of said oligomer lack the capability of interacting with the carboxy termini of gp36 polypeptides of bacteriophage T4.

6. An isolated polypeptide consisting essentially of a variant of the gp36 protein of bacteriophage T4, wherein

NSDOCID: <WO__9611947A1_IA>

said p lypeptide lacks the capability of interacting with the amino terminus of th P37 protein oligomer of bacteriophage T4.

7. An isolated polypeptide consisting essentially of a fusion protein between the gp36 and gp34 proteins of bacteriophage T4, wherein amino acid residues 1-73 of gp36 (SEQ ID NO:5) are fused in proper reading frame amino-terminal to amino acid residues 866-1289 of gp34 (SEQ ID NO:2).

8. An oligomer of the polypeptide of claim 7, wherein the amino termini of said dimer have the capability of interacting with the gp35 protein of bacteriophage T4.

9. An isolated polypeptide consisting essentially of a variant of the gp35 protein of bacteriophage T4, wherein said polypeptide forms an angle of less than about 125° when combined with the P34 and P36-P37 protein oligomers of bacteriophage T4, under conditions wherein the wild-type gp35 protein forms an angle of 137° when combined with said oligomers.

10. An isolated polypeptide consisting essentially of a variant of the gp35 protein of bacteriophage T4, wherein said polypeptide forms an angle of more than about 145° when combined with the P34 and P36-P37 protein oligomers of bacteriophage T4, under conditions wherein the wild-type gp35 protein forms an angle of 137° when combined with said oligomers.

11. An isolated polypeptide consisting essentially of a variant of the gp35 protein of bacteriophage T4, wherein the interaction of said polypeptide with the P34 protein oligomer of bacteriophag T4 is unstabl at temperatur s between about 40°C and about 60 C.

SDOCID: <WO 9611947A1 IA>

12. An isolated polypeptide oligomer consisting essentially of a variant of the P37 protein of bacteriophage T4, wherein the interaction of said oligomer with the P36 protein oligomer of bacteriophage T4 is unstable at temperatures between about 40°C and about 60°C.

13. An isolated polypeptide oligomer consisting essentially of a variant of the P37 protein of bacteriophage T4, wherein the carboxy-terminal domain of said oligomer is modified so as to confer the ability of the entire polypeptide to bind specifically to an immobilized ligand.

14. The polypeptide of claim 13, wherein said ligand is selected from the group consisting of biotin, immunoglobulin, or divalent metal ions.

15. A nanostructure comprising a plurality of fusion proteins, said fusion proteins comprising a first portion consisting of at least the first 10 N-terminal amino acids of a tail fiber protein fused via a peptide bond to a second portion consisting of at least the last 10 C-terminal amino acids of a second tail fiber protein, wherein the tail fiber proteins are selected from the group consisting of gp34, gp35, gp36, and gp37 proteins of a T-even-like bacteriophage, wherein the first and second tail fiber proteins are the same or different.

16. The nanostructure of claim 15, wherein the first and second tail fiber proteins are different.

17. The nanostructure of claim 15, which further comprises a molecule that can self-assemble into a dimer or trimer, fused to at least a 10 amino acid portion of a T-even-like tail fiber protein.

18. The nanostructure of claim 17, wherein the molecule has the structure of a leucine zipper.

NSDOCID: <WO__9611947A1_IA>

19. The nanostructure of claim 15, wherein said nanostructure comprises a linear one-dimensional rod.

20. The nanostructure of claim 15, wherein said nanostructure comprises a polygon.

21. The nanostructure of claim 15, wherein said nanostructure comprises a three-dimensional cage or solid.

22. The nanostructure of claim 15, wherein said nanostructure comprises a two-dimensional open or closed sheet.

23. An isolated fusion protein consisting essentially of a portion of a gp37 protein of a T-even-like bacteriophage consisting of at least the first 10-60 N-terminal amino acids of the gp37 protein fused to a second portion of a gp36 protein of a T-even-like bacteriophage consisting of at least the last 10-60 C-terminal amino acids of the gp36 protein.

24. An isolated fusion protein consisting essentially of a portion of a gp37 protein of a T-even-like bacteriophage consisting of at least the first 10 N-terminal amino acids of the gp37 protein fused to a second portion of a gp36 protein of a T-even-like bacteriophage consisting of at least the last 10 C-terminal amino acids of the gp36 protein.

25. An isolated fusion protein consisting essentially of a portion of a gp37 protein of a T-even-like bacteriophage consisting of at least the first 20 N-terminal amino acids of the gp37 protein fused to a second portion of a gp36 protein of a T-even-like bacteriophage consisting of at least the last 20 C-terminal amino acids of the gp36 protein.

26. An isolated fusion protein consisting essentially of a portion of a gp36 protein of a T-even-like bacteriophage consisting of at least the first 10-60 N-terminal amino acids of the gp36 protein fused to a second portion of a gp34 protein of a T-even-like bacteriophage consisting of at least the last 10-60 C-terminal amino acids of the gp34 protein.

27. An isolated protein comprising at least 20 contiguous amino acids of the gp37, gp36, or gp34 protein of a T-even-like bacteriophage, and lacking at least 5 amino acids of the amino- or carboxy-terminus of the protein.

28. An isolated DNA encoding the polypeptide of claim 1.

29. An isolated DNA encoding the polypeptide of claim 2.

30. An isolated DNA encoding the polypeptide of claim 4.

31. An isolated DNA encoding the polypeptide of claim 5.

32. An isolated DNA encoding the polypeptide of claim 6.

33. An isolated DNA encoding the polypeptide of claim 7.

34. An isolated DNA encoding the polypeptide of claim 9.

35. An isolated DNA encoding the polypeptide of claim 10.

NSDOCID: <WO___9611947A1_IA>

36. An isolated DNA encoding the polypeptide of claim 11.

37. An isolated DNA encoding the polypeptide of claim 12.

38. An isolated DNA encoding the polypeptide of claim 13.

39. An isolated DNA encoding the protein of claim 23.

40. An isolated DNA encoding the protein of claim 25.

41. An isolated DNA encoding the protein of claim 26.

42. An isolated DNA encoding the protein of claim 27.

43. A method for making a polygonal nanostructure comprising contacting the protein of claim 26 with purified gp35 proteins of a T-even-like bacteriophage.

44. A method for making a nanostructure comprising contacting a plurality of the proteins of claim 23 with each other.

45. A kit comprising in one or more containers the fusion protein of claim 23.

46. A kit comprising in one or more containers the fusion protein of claim 25.

47. A kit comprising in on or more containers the fusion pr tein f claim 26.

SNSDOCID: <WO 9611947A1 IA>

48. A kit comprising in ne or more containers the fusion protein of claim 26, and an is lated gp35 prot in of a T- ven-lik bacteriophag .

49. The protein of claim 23 wherein the T-even-like bacteriophage is T4.

50. The protein of claim 26 wherein the T-even-like bacteriophage is T4.

51. An isolated polypeptide consisting essentially of a variant of the gp36 protein of bacteriophage T4, wherein the interaction of said polypeptide with the P37 protein oligomer of bacteriophage T4 is unstable at temperatures between about 40°C and about 60°C.

52. An isolated polypeptide consisting essentially of a variant of the gp36 protein of bacteriophage T4, wherein the interaction of said polypeptide with the gp35 protein of bacteriophage T4 is unstable at temperatures between about 40°C and about 60°C.

53. An isolated polypeptide consisting essentially of a variant of the gp34 protein of bacteriophage T4, wherein the interaction of said polypeptide with the gp35 protein of bacteriophage T4 is unstable at temperatures between about 40°C and about 60°C.

!SDOCID: <WO__9611947A1_IA>

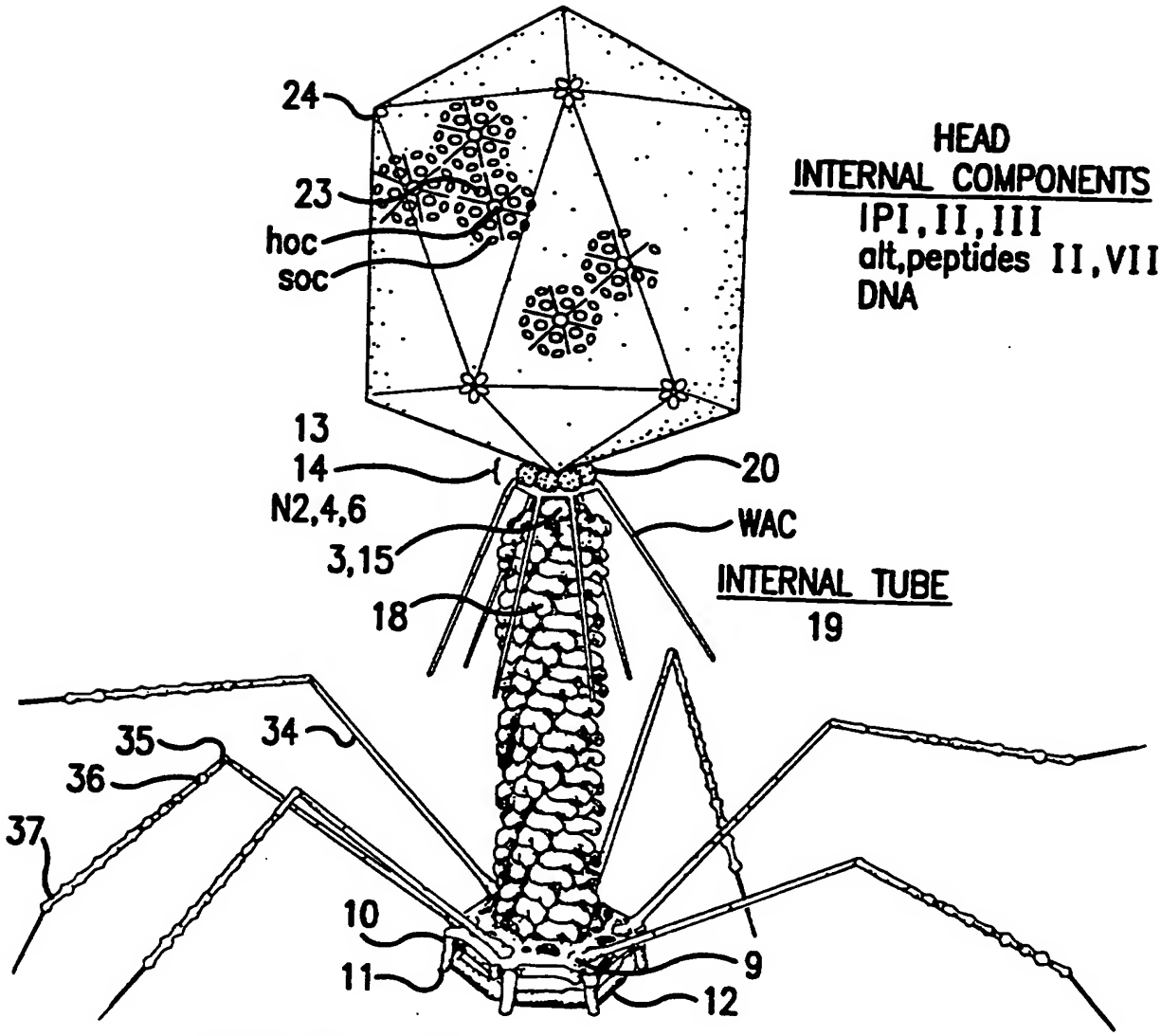


FIG. 1A

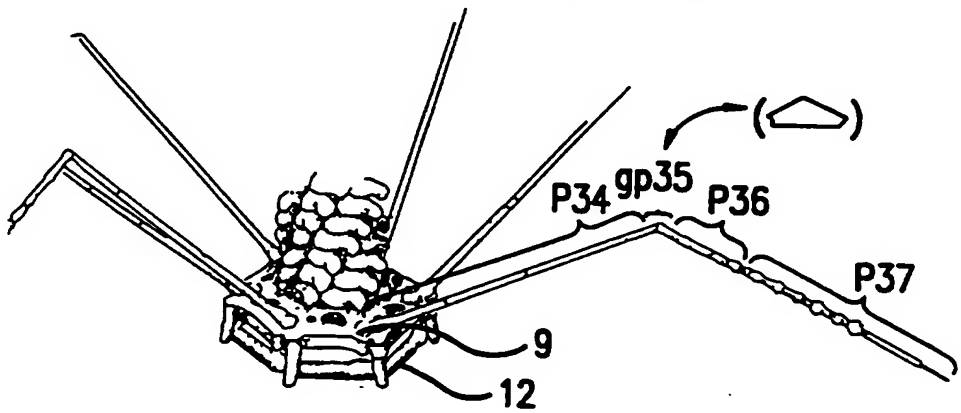


FIG. 1B

NSDOCID: <WO___9611947A1 IA>

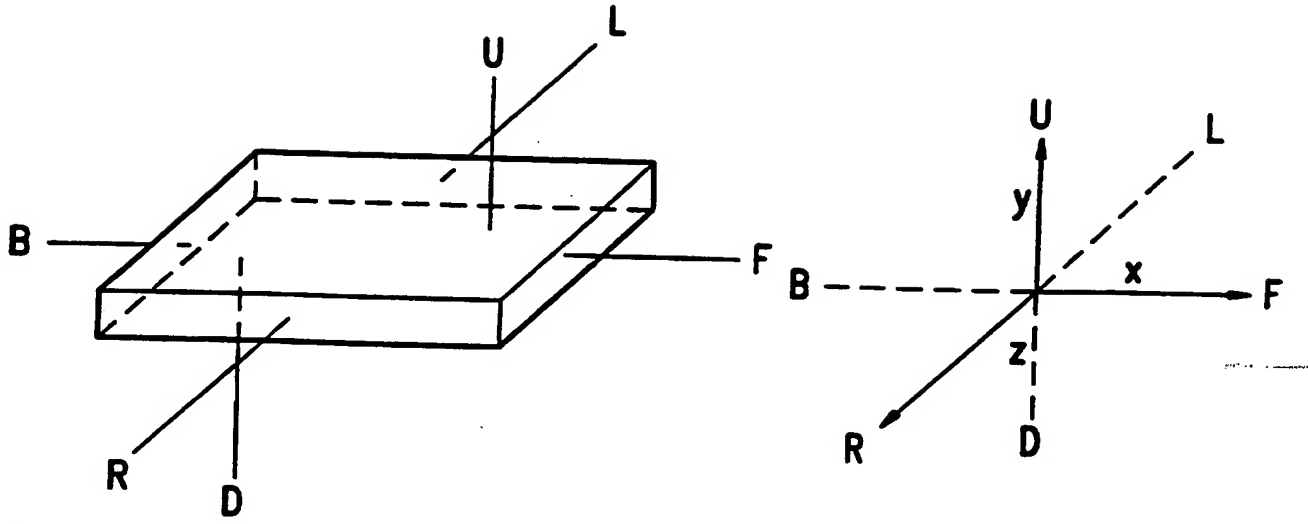


FIG.2

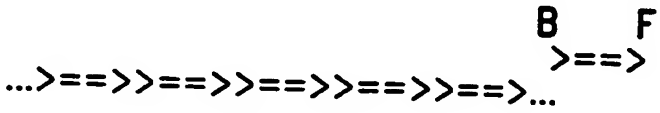


FIG.3A

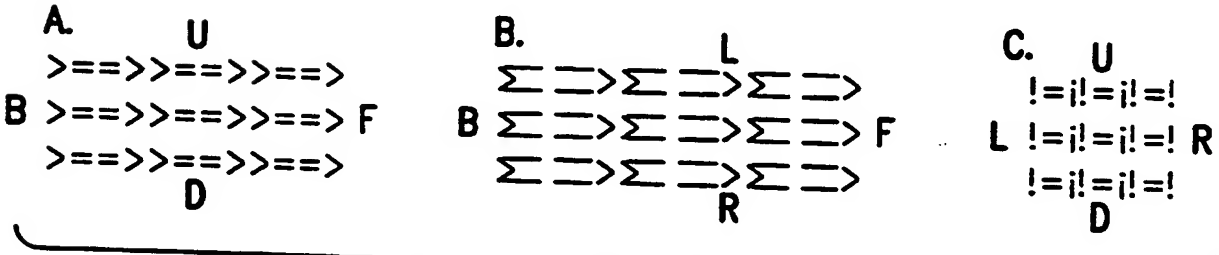


FIG.3B

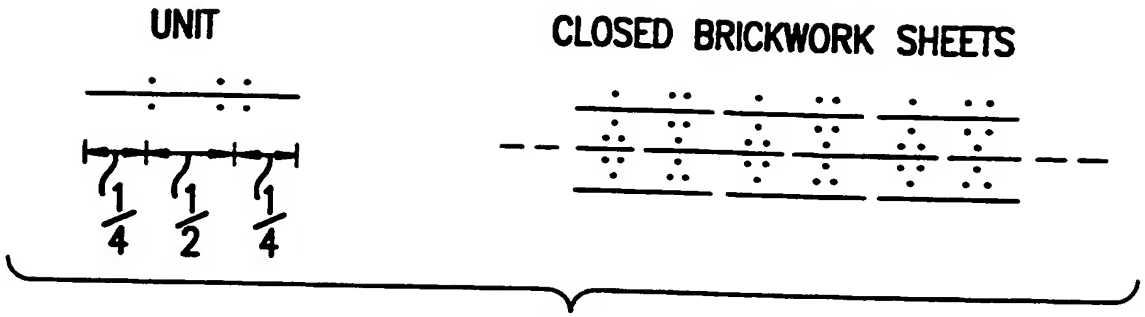

UNIT
CLOSED BRICKWORK SHEETS
1/4
1/2
1/4

FIG.3C

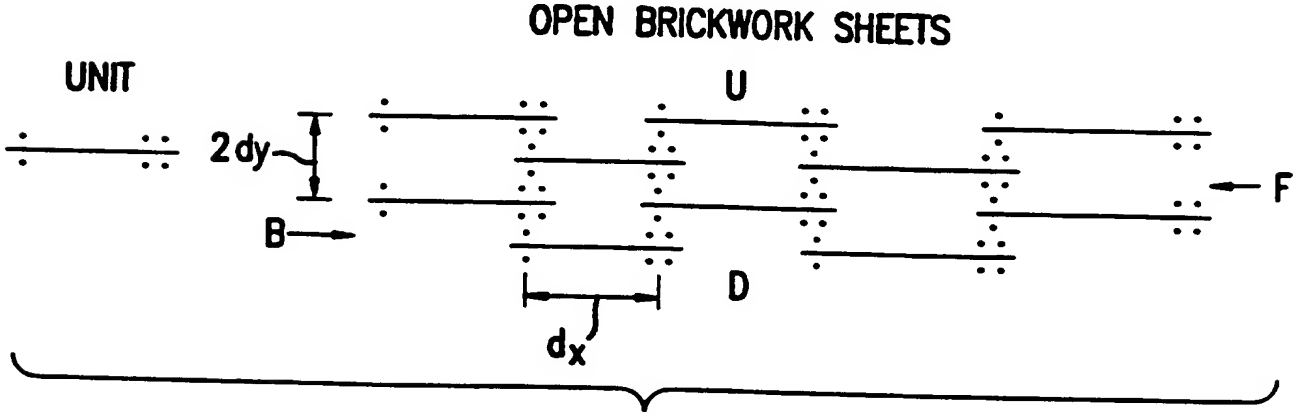

OPEN BRICKWORK SHEETS
UNIT
U
2dy
B
F
D
dx

FIG.3D

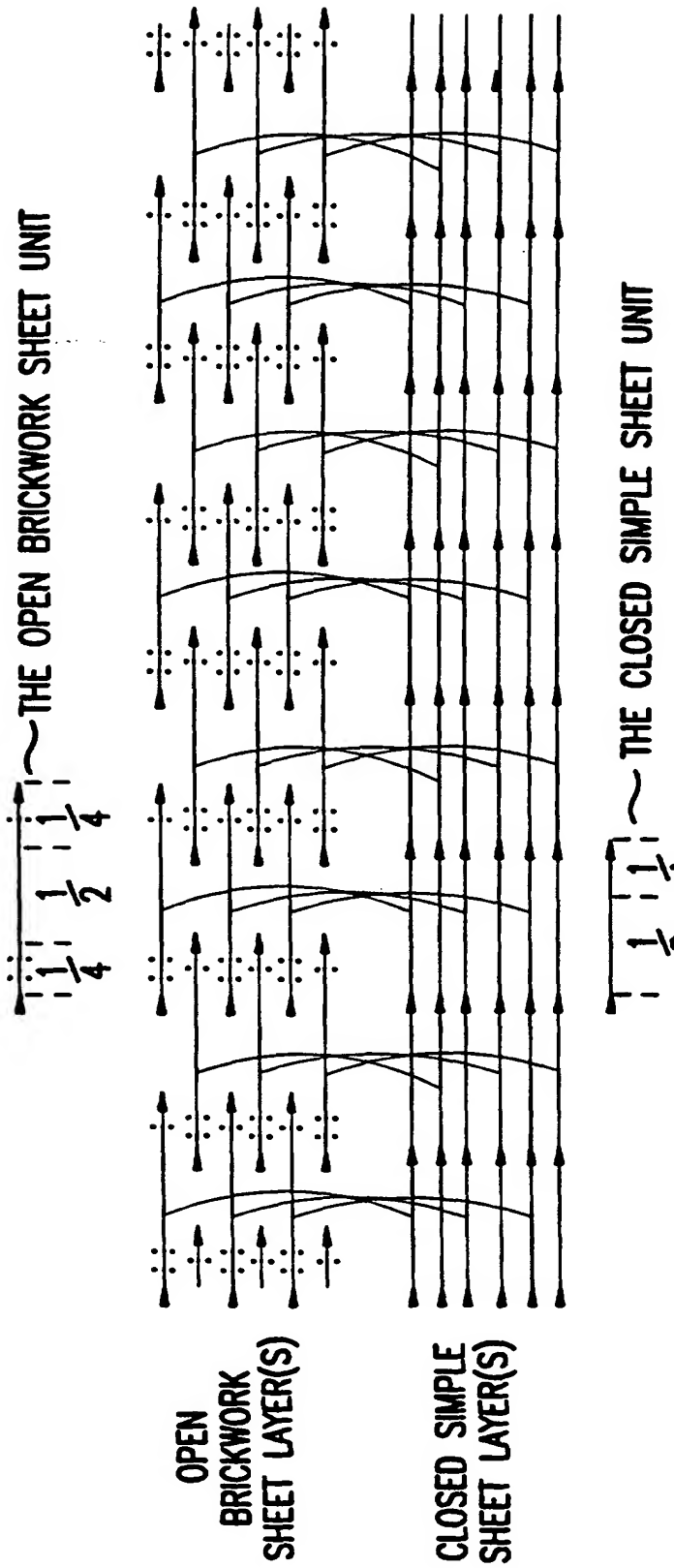

THE OPEN BRICKWORK SHEET UNIT
1/4
1/2
1 1/4
OPEN BRICKWORK SHEET LAYER(S)
CLOSED SIMPLE SHEET LAYER(S)
THE CLOSED SIMPLE SHEET UNIT
1/2
1 1/4

FIG.4

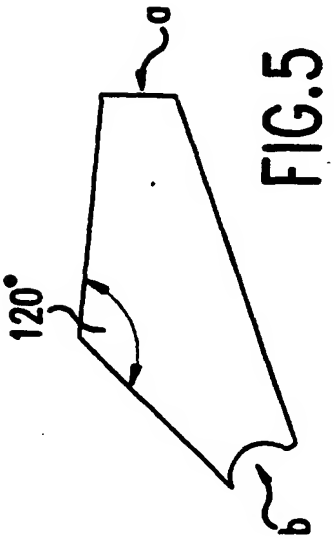

a
120°
b

FIG.5

	10	20	30	40	50	60	
1	TAGGAGCCCG	GGAGAATGGC	CGAGATTAAA	AGAGAATTCA	GAGCAGAAGA	TGGTCTGGAC	60
61	GCAGGTGGTG	ATAAAATAAT	CAACGTAGCT	TTAGCTGATC	GTACCGTAGG	AACTGACGGT	120
121	GTTAACGTTG	ATTACTTAAT	TCAAGAAAAC	ACAGTTCAAC	AGTATGATCC	AACTCGTGGA	180
181	TATTTAAAAG	ATTTTGTAAT	CATTTATGAT	AACCGCTTTT	GGGCTGCTAT	AAATGATATT	240
241	CCAAAACCAG	CAGGAGCTTT	TAATAGCGGA	CGCTGGAGAG	CATTACGTAC	CGATGCTAAC	300
301	TGGATTACGG	TTTCATCTGG	TTCATATCAA	TTAAAATCTG	GTGAAGCAAT	TTCGGTTAAC	360
361	ACCGCAGCTG	GAAATGACAT	CACGTTTACT	TTACCATCTT	CTCCAATTGA	TGGTGATACT	420
421	ATCGTTCTCC	AAGATATTGG	AGGAAAACCT	GGAGTTAACC	AAGTTTTAAT	TGTAGCTCCA	480
481	GTACAAAGTA	TTGTAAACTT	TAGAGGTGAA	CAGGTACGTT	CAGTACTAAT	GACTCATCCA	540
541	AAGTCACAGC	TAGTTTTAAT	TTTTAGTAAT	CGTCTGTGGC	AAATGTATGT	TGCTGATTAT	600
601	AGTAGAGAAG	CTATAGTTGT	AACACCAGCG	AATACTTATC	AAGCGCAATC	CAACGATTTT	660
661	ATCGTACGTA	GATTTACTTC	TGCTGCACCA	ATTAATGTCA	AACTTCCAAG	ATTTGCTAAT	720
721	CATGGCGATA	TTATTAATTT	CGTCGATTTA	GATAAACTAA	ATCCGCTTTA	TCATACAATT	780
781	GTTACTACAT	ACGATGAAAC	GACTTCAGTA	CAAGAAGTTG	GAACTCATTC	CATTGAAGGC	840
841	CGTACATCGA	TTGACGGTTT	CTTGATGTTT	GATGATAATG	AGAAATTATG	GAGACTGTTT	900
901	GACGGGGATA	GTAAAGCGCG	TTTACGTATC	ATAACGACTA	ATTCAAACAT	TCGTCCAAAT	960
961	GAAGAAGTTA	TGGTATTTGG	TGCGAATAAC	GGAACAACTC	AAACAATTGA	GCTTAAGCTT	1020
1021	CCAACTAATA	TTTCTGTTGG	TGATACTGTT	AAAATTTCCA	TGAATTACAT	GAGAAAAGGA	1080
1081	CAAACAGTTA	AAATCAAAGC	TGCTGATGAA	GATAAAATTG	CTTCTTCAGT	TCAATTGCTG	1140
1141	CAATTCCCAA	AACGCTCAGA	ATATCCACCT	GAAGCTGAAT	GGGTTACAGT	TCAAGAATTA	1200
1201	GTTTTTAACG	ATGAAACTAA	TTATGTTCCA	GTTTTGGAGC	TTGCTTACAT	AGAAGATTCT	1260
1261	GATGGAAAAT	ATTGGGTTGT	ACAGCAAAAC	GTTCCAACTG	TAGAAAGAGT	AGATTCTTTA	1320
1321	AATGATTCTA	CTAGAGCAAG	ATTAGGCGTA	ATTGCTTTAG	CTACACAAGC	TCAAGCTAAT	1380
1381	GTCGATTTAG	AAAATTCTCC	ACAAAAAGAA	TTAGCAATTA	CTCCAGAAAC	GTTAGCTAAT	1440
1441	CGTACTGCTA	CAGAAACTCG	CAGAGGTATT	GCAAGAATAG	CAACTACTGC	TCAAGTGAAT	1500
1501	CAGAACACCA	CATTCTCTTT	TGCTGATGAT	ATTATCATCA	CTCCTAAAAA	GCTGAATGAA	1560
1561	AGAACTGCTA	CAGAAACTCG	TAGAGGTGTC	GCAGAAATTG	CTACGCAGCA	AGAAACTAAT	1620
1621	GCAGGAACCG	ATGATACTAC	AATCATCACT	CCTAAAAAGC	TTCAAGCTCG	TCAAGGTTCT	1680
1681	GAATCATTAT	CTGGTATTGT	AACCTTTGTA	TCTACTGCAG	GTGCTACTCC	AGCTTCTAGC	1740
1741	CGTGAATTAA	ATGGTACGAA	TGTTTATAAT	AAAAACACTG	ATAATTTAGT	TGTTTCACCT	1800
1801	AAAGCTTTGG	ATCAGTATAA	AGCTACTCCA	ACACAGCAAG	GTGCAGTAAT	TTTAGCAGTT	1860
1861	GAAAGTGAAG	TAATTGCTGG	ACAAAGTCAG	CAAGGATGGG	CAAATGCTGT	TGTAACGCCA	1920
1921	GAAACGTTAC	ATAAAAAGAC	ATCAACTGAT	GGAAGAATTG	GTTTAATTGA	AATTGCTACG	1980
1981	CAAAGTGAAG	TTAATACAGG	AACTGATTAT	ACTCGTGCAG	TCACTCCTAA	AACTTTAAAT	2040
2041	GACCGTAGAG	CAACTGAAAG	TTTAAGTGGT	ATAGCTGAAA	TTGCTACACA	AGTTGAATTC	2100
2101	GACGCAGGCG	TCGACGATAC	TCGTATCTCT	ACACCATTAA	AAATTAAAAC	CAGATTTAAT	2160
2161	AGTACTGATC	GTACTTCTGT	TGTTGCTCTA	TCTGGATTAG	TTGAATCAGG	AACTCTCTGG	2220
2221	GACCATTATA	CACTTAATAT	TCTTGAAGCA	AATGAGACAC	AACGTGGTAC	ACTTCGTGTA	2280
2281	GCTACGCAGG	TCGAAGCTGC	TGCGGGAACA	TTAGATAATG	TTTTAATAAC	TCCTAAAAAG	2340

FIG.6A

SUBSTITUTE SHEET

SDOCID: <WO___9611947A1_IA>

2341 CTTTTAGGTA CTAAATCTAC TGAAGCGCAA GAGGGTGTTA TTAAAGTTGC AACTCAGTCT 2400
2401 GAAACTGTGA CTGGAACGTC AGCAAATACT GCTGTATCTC CAAAAAATTT AAAATGGATT 2460
2461 GCGCAGAGTG AACCTACTTG GGCAGCTACT ACTGCAATAA GAGGTTTTGT TAAAACTTCA 2520
2521 TCTGGTTCAA TTACATTCGT TGGTAATGAT ACAGTCGGTT CTACCCAAGA TTTAGAACTG 2580
2581 TATGAGAAAA ATAGCTATGC GGTATCACCA TATGAATTAA ACCGTGTATT AGCAAATTAT 2640
2641 TTGCCACTAA AAGCAAAAGC TGCTGATACA AATTTATTGG ATGGTCTAGA TTCATCTCAG 2700
2701 TTCATTCGTA GGGATATTGC ACAGACGGTT AATGGTTCAC TAACCTTAAC CCAACAAACG 2760
2761 AATCTGAGTG CCCCTCTTGT ATCATCTAGT ACTGGTGAAT TTGGTGGTTC ATTGGCCGCT 2820
2821 AATAGAACAT TTACCATCCG TAATACAGGA GCCCCGACTA GTATCGTTTT CGAAAAAGGT 2880
2881 CCTGCATCCG GGGCAAATCC TGCACAGTCA ATGAGTATTC GTCTATGGGG TAACCAATTT 2940
2941 GGCGGCGGTA GTGATACGAC CCGTTCGACA GTGTTTGAAG TTGGCGATGA CACATCTCAT 3000
3001 CACTTTTATT CTCAACGTAA TAAAGACGGT AATATAGCGT TTAACATTAA TGGTACTGTA 3060
3061 ATGCCAATAA ACATTAATGC TTCCGGTTTG ATGAATGTGA ATGGCACTGC AACATTCGGT 3120
3121 CGTTCAGTTA CAGCCAATGG TGAATTCATC AGCAAGTCTG CAAATGCTTT TAGAGCAATA 3180
3181 AACGGTGATT ACGGATTCTT TATTCGTAAT GATGCCTCTA ATACCTATTT TTTGCTCACT 3240
3241 GCAGCCGGTG ATCAGACTGG TGGTTTTAAT GGATTACGCC CATTATTAAT TAATAATCAA 3300
3301 TCCGGTCAGA TTACAATTGG TGAAGGCTTA ATCATTGCCA AAGGTGTTAC TATAAATTCA 3360
3361 GGCGGTTTAA CTGTTAACTC GAGAATTCGT TCTCAGGGTA CTAAAACATC TGATTTATAT 3420
3421 ACCCGTGCGC CAACATCTGA TACTGTAGGA TTCTGGTCAA TCGATATTAA TGATTCAGCC 3480
3481 ACTTATAACC AGTTCCCGGG TTATTTTAAA ATGGTTGAAA AAACTAATGA AGTGACTGGG 3540
3541 CTTCCATACT TAGAACGTGG CGAAGAAGTT AAATCTCCTG GTACACTGAC TCAGTTTGGT 3600
3601 AACACACTTG ATTCGCTTTA CCAAGATTCG ATTACTTATC CAACGACGCC AGAAGCGCGT 3660
3661 ACCACTCGCT GGACACGTAC ATGGCAGAAA ACCAAAAACT CTTGGTCAAG TTTTGTTCAG 3720
3721 GTATTTGACG GAGGTAACCC TCCTCAACCA TCTGATATCG GTGCTTTACC ATCTGATAAT 3780
3781 GCTACAATGG GGAATCTTAC TATTCGTGAT TTCTTGCGAA TTGGTAATGT TCGCATTGTT 3840
3841 CCTGACCCAG TGAATAAAAC GGTTAAATTT GAATGGGTTG AATAAGAGGT ATTATGGAAA 3900
3901 AATTTATGGC CGAGATTTGG ACAAGGATAT GTCCAAACGC CATTTTATCG GAAAGTAATT 3960
3961 CAGTAAGATA TAAAATAAGT ATAGCGGGTT CTTGCCCGCT TTCTACAGCA GGACCATCAT 4020
4021 ATGTTAAATT TCAGGATAAT CCTGTAGGAA GTCAAACATT TAGGCGCAGG CCTTCATTTA 4080
4081 AGAGTTTTTG ACCCTTCCAC CGGAGCATTA GTTGATAGTA AGTCATATGC TTTTTCGACT 4140
4141 TCAAATGATA CTACATCAGC TGCTTTTGTT AGTTTTCATG AATTCTTTGA CGAATAATCG 4200
4201 AATTGTTGCT ATATTAACTA GTGGAAAGGT TAATTTTCCT CCTGAAGTAG TATCTTGGTT 4260
4261 AAGAACCGCC GGAACGTCTG CCTTTCCATC TGATTCTATA TTGTCAAGAT TTGACGTATC 4320
4321 ATATGCTGCT TTTTATACTT CTTCTAAAAG AGCTATCGCA TTAGAGCATG TTAAACTGAG 4380
4381 TAATAGAAAA AGCACAGATG ATTATCAAAC TATTTTAGAT GTTGTATTTG ACAGTTTAGA 4440
4441 AGATGTAGGA GCTACCGGGT TTCCAAGAAG AACGTATGAA AGTGTTGAGC AATTCATGTC 4500
4501 GGCAGTTGGT GGAACTAATA ACGAAATTGC GAGATTGCCA ACTTCAGCTG CTATAAGTAA 4560
4561 ATTATCTGAT TATAATTTAA TTCCTGGAGA TGTTCTTTAT CTTAAAGCTC AGTTATATGC 4620
4621 TGATGCTGAT TTACTTGCTC TTGGAACTAC AAATATATCT ATCCGTTTTT ATAATGCATC 4680
4681 TAACGGATAT ATTTCTTCAA CACAAGCTGA ATTTACTGGG CAAGCTGGGT CATGGGAATT 4740

FIG.6B

BNSDOCID: <WO___9611947A1_IA>

4741	AAAGGAAGAT	TATGTAGTTG	TTCCAGAAAA	CGCAGTAGGA	TTTACGATAT	ACGCACAGAG	4800
4801	AACTGCACAA	GCTGGCCAAG	GTGGCATGAG	AAATTTAAGC	TTTTCTGAAG	TATCAAGAAA	4860
4861	TGGCGGCATT	TCGAAACCTG	CTGAATTTGG	CGTCAATGGT	ATTCGTGTTA	ATTATATCTG	4920
4921	CGAATCCGCT	TCACCTCCGG	ATATAATGGT	ACTTCCTACG	CAAGCATCGT	CTAAAACTGG	4980
4981	TAAAGTGTTT	GGGCAAGAAT	TTAGAGAAGT	TTAAATTGAG	GGACCCTTCG	GGTTCCCTTT	5040
5041	TTCTTTATAA	ATACTATTCA	AATAAAGGGG	CATACAATGG	CTGATTTAAA	AGTAGGTTCA	5100
5101	ACAACTGGAG	GCTCTGTCAT	TTGGCATCAA	GGAAATTTTC	CATTGAATCC	AGCCGGTGAC	5160
5161	GATGTACTCT	ATAAATCATT	TAAAATATAT	TCAGAATATA	ACAAACCACA	AGCTGCTGAT	5220
5221	AACGATTTCG	TTTCTAAAGC	TAATGGTGGT	ACTTATGCAT	CAAAGGTAAC	ATTTAACGCT	5280
5281	GGCATTCAAG	TCCCATATGC	TCCAAACATC	ATGAGCCCAT	GCGGGATTTA	TGGGGGTAAC	5340
5341	GGTGATGGTG	CTACTTTTGA	TAAAGCAAAT	ATCGATATTG	TTTCATGGTA	TGGCGTAGGA	5400
5401	TTTAAATCGT	CATTTGGTTC	AACAGGCCGA	ACTGTTGTAA	TTAATACACG	CAATGGTGAT	5460
5461	ATTAACACAA	AAGGTGTTGT	GTCGGCAGCT	GGTCAAGTAA	GAAGTGGTGC	GGCTGCTCCT	5520
5521	ATAGCAGCGA	ATGACCTTAC	TAGAAAGGAC	TATGTTGATG	GAGCAATAAA	TACTGTTACT	5580
5581	GCAAATGCAA	ACTCTAGGGT	GCTACGGTCT	GGTGACACCA	TGACAGGTAA	TTTAACAGCG	5640
5641	CCAAACTTTT	TCTCGCAGAA	TCCTGCATCT	CAACCCTCAC	ACGTTCCACG	ATTTGACCAA	5700
5701	ATCGTAATTA	AGGATTCTGT	TCAAGATTTC	GGCTATTATT	AAGAGGACTT	ATGGCTACTT	5760
5761	TAAAACAAAT	ACAATTTAAA	AGAAGCAAAA	TCGCAGGAAC	ACGTCCTGCT	GCTTCAGTAT	5820
5821	TAGCCGAAGG	TGAATTGGCT	ATAAACTTAA	AAGATAGAAC	AATTTTTACT	AAAGATGATT	5880
5881	CAGGAAATAT	CATCGATCTA	GGTTTTGCTA	AAGGCGGGCA	AGTTGATGGC	AACGTTACTA	5940
5941	TTAACGGACT	TTTGAGATTA	AATGCCGATT	ATGTACAAAC	AGGTGGAATG	ACTGTAAACG	6000
6001	GACCCATTGG	TTCTACTGAT	GGCGTCACTG	GAAAAATTTT	CAGATCTACA	CAGGGTTCAT	6060
6061	TTTATGCAAG	AGCAACAAAC	GATACTTCAA	ATGCCCATTT	ATGGTTTGAA	AATGCCGATG	6120
6121	GCACTGAACG	TGGCGTTATA	TATGCTCGCC	CTCAAACTAC	AACTGACGGT	GAAATACGCC	6180
6181	TTAGGGTTAG	ACAAGGAACA	GGAAGCACTG	CCAACAGTGA	ATTCTATTTC	CGCTCTATAA	6240
6241	ATGGAGGCGA	ATTTCAGGCT	AACCGTATTT	TAGCATCAGA	TTCGTTAGTA	ACAAAACGCA	6300
6301	TTGCGGTTGA	TACCGTTATT	CATGATGCCA	AAGCATTTGG	ACAATATGAT	TCTCACTCTT	6360
6361	TGGTTAATTA	TGTTTATCCT	GGAACCGGTG	AAACAAATGG	TGTAAACTAT	CTTCGTAAAG	6420
6421	TTCGCGCTAA	GTCCGGTGGT	ACAATTTATC	ATGAAATTGT	TACTGCACAA	ACAGGCCTGG	6480
6481	CTGATGAAGT	TTCTTGGTGG	TCTGGTGATA	CACCAGTATT	TAAACTATAC	GGTATTCGTG	6540
6541	ACGATGGCAG	AATGATTATC	CGTAATAGCC	TTGCATTAGG	TACATTCACT	ACAAATTTCC	6600
6601	CGTCTAGTGA	TTATGGCAAC	GTCGGTGTAA	TGGGCGATAA	GTATCTTGTT	CTCGGCGACA	6660
6661	CTGTAACTGG	CTTGTCATAC	AAAAAAACTG	GTGTATTTGA	TCTAGTTGGC	GGTGGATATT	6720
6721	CTGTTGCTTC	TATTACTCCT	GACAGTTTCC	GTAGTACTCG	TAAAGGTATA	TTTGGTCGTT	6780
6781	CTGAGGACCA	AGGCGCAACT	TGGATAATGC	CTGGTACAAA	TGCTGCTCTC	TTGTCTGTTC	6840
6841	AAACACAAGC	TGATAATAAC	AATGCTGGAG	ACGGACAAAC	CCATATCGGG	TACAATGCTG	6900
6901	GCGGTAAAAT	GAACCACTAT	TTCCGTGGTA	CAGGTCAGAT	GAATATCAAT	ACCCAACAAG	6960
6961	GTATGGAAAT	TAACCCGGGT	ATTTTGAAAT	TGGTAACTGG	CTCTAATAAT	GTACAATTTT	7020
7021	ACGCTGACGG	AACTATTTCT	TCCATTCAAC	CTATTAAATT	AGATAACGAG	ATATTTTTAA	7080
7081	CTAAATCTAA	TAATACTGCG	GGTCTTAAAT	TTGGAGCTCC	TAGCCAAGTT	GATGGCACAA	7140

FIG.6C

SNSDOCID: <WO___9611947A1 IA>

```
7141  GGACTATCCA ATGGAACGGT GGTACTCGCG  AAGGACAGAA TAAAAACTAT GTGATTATTA  7200
7201  AAGCATGGGG TAACTCATTT AATGCCACTG  GTGATAGATC TCGCGAAACG GTTTTCCAAG  7260
7261  TATCAGATAG TCAAGGATAT TATTTTTATG  CTCATCGTAA AGCTCCAACC GGCGACGAAA  7320
7321  CTATTGGACG TATTGAAGCT CAATTTGCTG  GGGATGTTTA TGCTAAAGGT ATTATTGCCA  7380
7381  ACGGAAATTT TAGAGTTGTT GGGTCAAGCG  CTTTAGCCGG CAATGTTACT ATGTCTAACG  7440
7441  GTTTGTTTGT CCAAGGTGGT TCTTCTATTA  CTGGACAAGT TAAAATTGGC GGAACAGCAA  7500
7501  ACGCACTGAG AATTTGGAAC GCTGAATATG  GTGCTATTTT CCGTCGTTCG GAAAGTAACT  7560
7561  TTTATATTAT TCCAACCAAT CAAAATGAAG  GAGAAAGTGG AGACATTCAC AGCTCTTTGA  7620
7621  GACCTGTGAG AATAGGATTA AACGATGGCA  TGGTTGGGTT AGGAAGAGAT TCTTTTATAG  7680
7681  TAGATCAAAA TAATGCTTTA ACTACGATAA  ACAGTAACTC TCGCATTAAT GCCAACTTTA  7740
7741  GAATGCAATT GGGGCAGTCG GCATACATTG  ATGCAGAATG TACTGATGCT GTTCGCCCGG  7800
7801  CGGGTGCAGG TTCATTTGCT TCCCAGAATA  ATGAAGACGT CCGTGCGCCG TTCTATATGA  7860
7861  ATATTGATAG AACTGATGCT AGTGCATATG  TTCCTATTTT GAAACAACGT TATGTTCAAG  7920
7921  GCAATGGCTG CTATTCATTA GGGACTTTAA  TTAATAATGG TAATTTCCGA GTTCATTACC  7980
7981  ATGGCGGCGG AGATAACGGT TCTACAGGTC  CACAGACTGC TGATTTTGGA TGGGAATTTA  8040
8041  TTAAAAACGG TGATTTTATT TCACCTCGCG  ATTTAATAGC AGGCAAAGTC AGATTTGATA  8100
8101  GAACTGGTAA TATCACTGGT GGTTCTGGTA  ATTTTGCTAA CTTAAACAGT ACAATTGAAT  8160
8161  CACTTAAAAC TGATATCATG TCGAGTTACC  CAATTGGTGC TCCGATTCCT TGGCCGAGTG  8220
8221  ATTCAGTTCC TGCTGGATTT GCTTTGATGG  AAGGTCAGAC CTTTGATAAG TCCGCATATC  8280
8281  CAAAGTTAGC TGTTGCATAT CCTAGCGGTG  TTATTCCAGA TATGCGCGGG CAAACTATCA  8340
8341  AGGGTAAACC AAGTGGTCGT GCTGTTTTGA  GCGCTGAGGC AGATGGTGTT AAGGCTCATA  8400
8401  GCCATAGTGC ATCGGCTTCA AGTACTGACT  TAGGTACTAA AACCACATCA AGCTTTGACT  8460
8461  ATGGTACGAA GGGAACTAAC AGTACGGGTG  GACACACTCA CTCTGGTAGT GGTTCTACTA  8520
8521  GCACAAATGG TGAGCACAGC CACTACATCG  AGGCATGGAA TGGTACTGGT GTAGGTGGTA  8580
8581  ATAAGATGTC ATCATATGCC ATATCATACA  GGGCGGGTGG GAGTAACACT AATGCAGCAG  8640
8641  GGAACCACAG TCACACTTTC TCTTTTGGGA  CTAGCAGTGC TGGCGACCAT TCCCACTCTG  8700
8701  TAGGTATTGG TGCTCATACC CACACGGTAG  CAATTGGATC ACATGGTCAT ACTATCACTG  8760
8761  TAAATAGTAC AGGTAATACA GAAAACACGG  TTAAAAACAT TGCTTTTAAC TATATCGTTC  8820
8821  GTTTAGCATA AGGAGAGGGG CTTCGGCCCT  TCTAA                             8855
          |  10     |  20     |  30        |  40     |  50     |  60
```

FIG.6D

BNSDOCID: <WO___9611947A1_IA>

SUBSTITUTE SHEET (RULE 26)

RSDOCID: <WO 9611947A1_IA>

```
1   TAGGAGCCCGGGAGA ATG GCC GAG ATT AAA AGA GAA TTC AGA GCA GAA GAT GGT CTG GAC GCA    63
1                   M  A  E  I  K  R  E  F  R  A  E  D  G  L  D  A                   16

64  GGT GGT GAT AAA ATA ATC AAC GTA GCT TTA GCT GAT CGT ACC GTA GGA ACT GAC GGT GTT   123
17  G  G  D  K  I  I  N  V  A  L  A  D  R  T  V  G  T  D  G  V                      36

124 AAC GTT GAT TAC TTA ATT CAA GAA AAC ACA GTT CAA CAG TAT GAT CCA ACT CGT GGA TAT   183
37  N  V  D  Y  L  I  Q  E  N  T  V  Q  Q  Y  D  P  T  R  G  Y                      56

184 TTA AAA GAT TTT GTA ATC ATT TAT GAT AAC CGC TTT TGG GCT GCT ATA AAT GAT ATT CCA   243
57  L  K  D  F  V  I  I  Y  D  N  R  F  W  A  A  I  N  D  I  P                      76

244 AAA CCA GCA GGA GCT TTT AAT AGC GGA CGC TGG AGA GCA TTA CGT ACC GAT GCT AAC TGG   303
77  K  P  A  G  A  F  N  S  G  R  W  R  A  L  R  T  D  A  N  W                      96

304 ATT ACG GTT TCA TCT GGT TCA TAT CAA TTA AAA TCT GGT GAA GCA ATT TCG GTT AAC ACC   363
97  I  T  V  S  S  G  S  Y  Q  L  K  S  G  E  A  I  S  V  N  T                     116

364 GCA GCT GGA AAT GAC ATC ACG TTT ACT TTA CCA TCT TCT CCA ATT GAT GGT GAT ACT ATC   423
117 A  A  G  N  D  I  T  F  T  L  P  S  S  P  I  D  G  D  T  I                     136

424 GTT CTC CAA GAT ATT GGA GGA AAA CCT GGA GTT AAC CAA GTT TTA ATT GTA GCT CCA GTA   483
137 V  L  Q  D  I  G  G  K  P  G  V  N  Q  V  L  I  V  A  P  V                     156

484 CAA AGT ATT GTA AAC TTT AGA GGT GAA CAG GTA CGT TCA GTA CTA ATG ACT CAT CCA AAG   543
157 Q  S  I  V  N  F  R  G  E  Q  V  R  S  V  L  M  T  H  P  K                     176
```

FIG.7A

SUBSTITUTE SHEET (RULE 26)

<SDOCID: <WO___9611947A1_IA>

```
544  TCA CAG CTA GTT TTA ATT TTT AGT AAT CGT CTG TGG CAA ATG TAT GTT GCT GAT TAT AGT  603
177  S   Q   L   V   L   I   F   S   N   R   L   W   Q   M   Y   V   A   D   Y   S    196

604  AGA GAA GCT ATA GTT GTA ACA CCA GCG AAT ACT TAT CAA GCG CAA TCC AAC GAT TTT ATC  663
197  R   E   A   I   V   V   T   P   A   N   T   Y   Q   A   Q   S   N   D   F   I    216

664  GTA CGT AGA TTT ACT TCT GCT GCA CCA ATT AAT GTC AAA CTT CCA AGA TTT GCT AAT CAT  723
217  V   R   R   F   T   S   A   A   P   I   N   V   K   L   P   R   F   A   N   H    236

724  GGC GAT ATT ATT AAT TTC GTC GAT TTA GAT AAA CTA AAT CCG CTT TAT CAT ACA ATT GTT  783
237  G   D   I   I   N   F   V   D   L   D   K   L   N   P   L   Y   H   T   I   V    256

784  ACT ACA TAC GAT GAA ACG ACT TCA GTA CAA GAA GTT GGA ACT CAT TCC ATT GAA GGC CGT  843
257  T   T   Y   D   E   T   T   S   V   Q   E   V   G   T   H   S   I   E   G   R    276

844  ACA TCG ATT GAC GGT TTC TTG ATG TTT GAT GAT AAT GAG AAA TTA TGG AGA CTG TTT GAC  903
277  T   S   I   D   G   F   L   M   F   D   D   N   E   K   L   W   R   L   F   D    296

904  GGG GAT AGT AAA GCG CGT TTA CGT ATC ATA ACG ACT AAT TCA AAC ATT CGT CCA AAT GAA  963
297  G   D   S   K   A   R   L   R   I   I   T   T   N   S   N   I   R   P   N   E    316

964  GAA GTT ATG GTA TTT GGT GCG AAT AAC GGA ACA ACT CAA ACA ATT GAG CTT AAG CTT CCA 1023
317  E   V   M   V   F   G   A   N   N   G   T   T   Q   T   I   E   L   K   L   P    336

1024 ACT AAT ATT TCT GTT GGT GAT ACT GTT AAA ATT TCC ATG AAT TAC ATG AGA AAA GGA CAA 1083
337  T   N   I   S   V   G   D   T   V   K   I   S   M   N   Y   M   R   K   G   Q    356
```

FIG.7B

3NSDOCID: <WO__9611947A1_IA>


```
1084  ACA GTT AAA ATC AAA GCT GCT GAT GAA GAT AAA ATT GCT TCT TCA GTT CAA TTG CTG CAA  1143
 357  T   V   K   I   K   A   A   D   E   D   K   I   A   S   S   V   Q   L   L   Q    376

1144  TTC CCA AAA CGC TCA GAA TAT CCA CCT GAA GCT GAA TGG GTT ACA GTT CAA GAA TTA GTT  1203
 377  F   P   K   R   S   E   Y   P   P   E   A   E   W   V   T   V   Q   E   L   V    396

1204  TTT AAC GAT GAA ACT AAT TAT GTT CCA GTT TTG GAG CTT GCT TAC ATA GAA GAT TCT GAT  1263
 397  F   N   D   E   T   N   Y   V   P   V   L   E   L   A   Y   I   E   D   S   D    416

1264  GGA AAA TAT TGG GTT GTA CAG CAA AAC GTT CCA ACT GTA GAA AGA GTA GAT TCT TTA AAT  1323
 417  G   K   Y   W   V   V   Q   Q   N   V   P   T   V   E   R   V   D   S   L   N    436

1324  GAT TCT ACT AGA GCA AGA TTA GGC GTA ATT GCT TTA GCT ACA CAA GCT CAA GCT AAT GTC  1383
 437  D   S   T   R   A   R   L   G   V   I   A   L   A   T   Q   A   Q   A   N   V    456

1384  GAT TTA GAA AAT TCT CCA CAA AAA GAA TTA GCA ATT ACT CCA GAA ACG TTA GCT AAT CGT  1443
 457  D   L   E   N   S   P   Q   K   E   L   A   I   T   P   E   T   L   A   N   R    476

1444  ACT GCT ACA GAA ACT CGC AGA GGT ATT GCA AGA ATA GCA ACT ACT GCT CAA GTG AAT CAG  1503
 477  T   A   T   E   T   R   R   G   I   A   R   I   A   T   T   A   Q   V   N   Q    496

1504  AAC ACC ACA TTC TCT TTT GCT GAT GAT ATT ATC ATC ACT CCT AAA AAG CTG AAT GAA AGA  1563
 497  N   T   T   F   S   F   A   D   D   I   I   I   T   P   K   K   L   N   E   R    516

1564  ACT GCT ACA GAA ACT CGT AGA GGT GTC GCA GAA ATT GCT ACG CAG CAA GAA ACT AAT GCA  1623
 517  T   A   T   E   T   R   R   G   V   A   E   I   A   T   Q   Q   E   T   N   A    536
```

FIG.7C

BNSDOCID: <WO 9611947A1_IA>

SUBSTITUTE SHEET (RULE 26)

```
1624  GGA ACC GAT GAT ACT ACA ATC ATC ACT CCT AAA AAG CTT CAA GCT CGT CAA GGT TCT GAA  1683
 537   G   T   D   D   T   T   I   I   T   P   K   K   L   Q   A   R   Q   G   S   E    556

1684  TCA TTA TCT GGT ATT GTA ACC TTT GTA TCT ACT GCA GGT GCT ACT CCA GCT TCT AGC CGT  1743
 557   S   L   S   G   I   V   T   F   V   S   T   A   G   A   T   P   A   S   S   R    576

1744  GAA TTA AAT GGT ACG AAT GTT TAT AAT AAA AAC ACT GAT AAT TTA GTT GTT TCA CCT AAA  1803
 577   E   L   N   G   T   N   V   Y   N   K   N   T   D   N   L   V   V   S   P   K    596

1804  GCT TTG GAT CAG TAT AAA GCT ACT CCA ACA CAG CAA GGT GCA GTA ATT TTA GCA GTT GAA  1863
 597   A   L   D   Q   Y   K   A   T   P   T   Q   Q   G   A   V   I   L   A   V   E    616

1864  AGT GAA GTA ATT GCT GGA CAA AGT CAG CAA GGA TGG GCA AAT GCT GTT GTA ACG CCA GAA  1923
 617   S   E   V   I   A   G   Q   S   Q   Q   G   W   A   N   A   V   V   T   P   E    636

1924  ACG TTA CAT AAA AAG ACA TCA ACT GAT GGA AGA ATT GGT TTA ATT GAA ATT GCT ACG CAA  1983
 637   T   L   H   K   K   T   S   T   D   G   R   I   G   L   I   E   I   A   T   Q    656

1984  AGT GAA GTT AAT ACA GGA ACT GAT TAT ACT CGT GCA GTC ACT CCT AAA ACT TTA AAT GAC  2043
 657   S   E   V   N   T   G   T   D   Y   T   R   A   V   T   P   K   T   L   N   D    676

2044  CGT AGA GCA ACT GAA AGT TTA AGT GGT ATA GCT GAA ATT GCT ACA CAA GTT GAA TTC GAC  2103
 677   R   R   A   T   E   S   L   S   G   I   A   E   I   A   T   Q   V   E   F   D    696

2104  GCA GGC GTC GAC GAT ACT CGT ATC TCT ACA CCA TTA AAA ATT AAA ACC AGA TTT AAT AGT  2163
 697   A   G   V   D   D   T   R   I   S   T   P   L   K   I   K   T   R   F   N   S    716
```

FIG.7D

ISDOCID: <WO__9611947A1_IA>

SUBSTITUTE SHEET (RULE 26)

```
2164  ACT GAT CGT ACT TCT GTT GTT GCT CTA TCT GGA TTA GTT GAA TCA GGA ACT CTC TGG GAC  2223
 717  T   D   R   T   S   V   V   A   L   S   G   L   V   E   S   G   T   L   W   D    736

2224  CAT TAT ACA CTT AAT ATT CTT GAA GCA AAT GAG ACA CAA CGT GGT ACA CTT CGT GTA GCT  2283
 737  H   Y   T   L   N   I   L   E   A   N   E   T   Q   R   G   T   L   R   V   A    756

2284  ACG CAG GTC GAA GCT GCT GCG GGA ACA TTA GAT AAT GTT TTA ATA ACT CCT AAA AAG CTT  2343
 757  T   Q   V   E   A   A   A   G   T   L   D   N   V   L   I   T   P   K   K   L    776

2344  TTA GGT ACT AAA TCT ACT GAA GCG CAA GAG GGT GTT ATT AAA GTT GCA ACT CAG TCT GAA  2403
 777  L   G   T   K   S   T   E   A   Q   E   G   V   I   K   V   A   T   Q   S   E    796

2404  ACT GTG ACT GGA ACG TCA GCA AAT ACT GCT GTA TCT CCA AAA AAT TTA AAA TGG ATT GCG  2463
 797  T   V   T   G   T   S   A   N   T   A   V   S   P   K   N   L   K   W   I   A    816

2464  CAG AGT GAA CCT ACT TGG GCA GCT ACT ACT GCA ATA AGA GGT TTT GTT AAA ACT TCA TCT  2523
 817  Q   S   E   P   T   W   A   A   T   T   A   I   R   G   F   V   K   T   S   S    836

2524  GGT TCA ATT ACA TTC GTT GGT AAT GAT ACA GTC GGT TCT ACC CAA GAT TTA GAA CTG TAT  2583
 837  G   S   I   T   F   V   G   N   D   T   V   G   S   T   Q   D   L   E   L   Y    856

2584  GAG AAA AAT AGC TAT GCG GTA TCA CCA TAT GAA TTA AAC CGT GTA TTA GCA AAT TAT TTG  2643
 857  E   K   N   S   Y   A   V   S   P   Y   E   L   N   R   V   L   A   N   Y   L    876

2644  CCA CTA AAA GCA AAA GCT GCT GAT ACA AAT TTA TTG GAT GGT CTA GAT TCA TCT CAG TTC  2703
 877  P   L   K   A   K   A   A   D   T   N   L   L   D   G   L   D   S   S   Q   F    896
```

FIG.7E

SUBSTITUTE SHEET (RULE 26)

<WO__9611947A1_IA>

```
2704  ATT CGT AGG GAT ATT GCA CAG ACG GTT AAT GGT TCA CTA ACC TTA ACC CAA CAA ACG AAT  2763
 897  I   R   R   D   I   A   Q   T   V   N   G   S   L   T   L   T   Q   Q   T   N    916

2764  CTG AGT GCC CCT CTT GTA TCA TCT AGT ACT GGT GAA TTT GGT GGT TCA TTG GCC GCT AAT  2823
 917  L   S   A   P   L   V   S   S   S   T   G   E   F   G   G   S   L   A   A   N    936

2824  AGA ACA TTT ACC ATC CGT AAT ACA GGA GCC CCG ACT AGT ATC GTT TTC GAA AAA GGT CCT  2883
 937  R   T   F   T   I   R   N   T   G   A   P   T   S   I   V   F   E   K   G   P    956

2884  GCA TCC GGG GCA AAT CCT GCA CAG TCA ATG AGT ATT CGT GTA TGG GGT AAC CAA TTT GGC  2943
 957  A   S   G   A   N   P   A   Q   S   M   S   I   R   V   W   G   N   Q   F   G    976

2944  GGC GGT AGT GAT ACG ACC CGT TCG ACA GTG TTT GAA GTT GGC GAT GAC ACA TCT CAT CAC  3003
 977  G   G   S   D   T   T   R   S   T   V   F   E   V   G   D   D   T   S   H   H    996

3004  TTT TAT TCT CAA CGT AAT AAA GAC GGT AAT ATA GCG TTT AAC ATT AAT GGT ACT GTA ATG  3063
 997  F   Y   S   Q   R   N   K   D   G   N   I   A   F   N   I   N   G   T   V   M   1016

3064  CCA ATA AAC ATT AAT GCT TCC GGT TTG ATG AAT GTG AAT GGC ACT GCA ACA TTC GGT CGT  3123
1017  P   I   N   I   N   A   S   G   L   M   N   V   N   G   T   A   T   F   G   R   1036

3124  TCA GTT ACA GCC AAT GGT GAA TTC ATC AGC AAG TCT GCA AAT GCT TTT AGA GCA ATA AAC  3183
1037  S   V   T   A   N   G   E   F   I   S   K   S   A   N   A   F   R   A   I   N   1056

3184  GGT GAT TAC GGA TTC TTT ATT CGT AAT GAT GCC TCT AAT ACC TAT TTT TTG CTC ACT GCA  3243
1057  G   D   Y   G   F   F   I   R   N   D   A   S   N   T   Y   F   L   L   T   A   1076
```

FIG.7F

SUBSTITUTE SHEET (RULE 26)

<SDOCID: <WO 9611947A1 IA>

```
3244  GCC GGT GAT CAG ACT GGT GGT TTT AAT GGA TTA CGC CCA TTA TTA ATT AAT AAT CAA TCC  3303
1077  A  G  D  Q  T  G  G  F  N  G  L  R  P  L  L  I  N  N  Q  S                    1096

3304  GGT CAG ATT ACA ATT GGT GAA GGC TTA ATC ATT GCC AAA GGT GTT ACT ATA AAT TCA GGC  3363
1097  G  Q  I  T  I  G  E  G  L  I  I  A  K  G  V  T  I  N  S  G                    1116

3364  GGT TTA ACT GTT AAC TCG AGA ATT CGT TCT CAG GGT ACT AAA ACA TCT GAT TTA TAT ACC  3423
1117  G  L  T  V  N  S  R  I  R  S  Q  G  T  K  T  S  D  L  Y  T                    1136

3424  CGT GCG CCA ACA TCT GAT ACT GTA GGA TTC TGG TCA ATC GAT ATT AAT GAT TCA GCC ACT  3483
1137  R  A  P  T  S  D  T  V  G  F  W  S  I  D  I  N  D  S  A  T                    1156

3484  TAT AAC CAG TTC CCG GGT TAT TTT AAA ATG GTT GAA AAA ACT AAT GAA GTG ACT GGG CTT  3543
1157  Y  N  Q  F  P  G  Y  F  K  M  V  E  K  T  N  E  V  T  G  L                    1176

3544  CCA TAC TTA GAA CGT GGC GAA GAA GTT AAA TCT CCT GGT ACA CTG ACT CAG TTT GGT AAC  3603
1177  P  Y  L  E  R  G  E  E  V  K  S  P  G  T  L  T  Q  F  G  N                    1196

3604  ACA CTT GAT TCG CTT TAC CAA GAT TGG ATT ACT TAT CCA ACG ACG CCA GAA GCG CGT ACC  3663
1197  T  L  D  S  L  Y  Q  D  W  I  T  Y  P  T  T  P  E  A  R  T                    1216

3664  ACT CGC TGG ACA CGT ACA TGG CAG AAA ACC AAA AAC TCT TGG TCA AGT TTT GTT CAG GTA  3723
1217  T  R  W  T  R  T  W  Q  K  T  K  N  S  W  S  S  F  V  Q  V                    1236

3724  TTT GAC GGA GGT AAC CCT CCT CAA CCA TCT GAT ATC GGT GCT TTA CCA TCT GAT AAT GCT  3783
1237  F  D  G  G  N  P  P  Q  P  S  D  I  G  A  L  P  S  D  N  A                    1256
```

FIG.7G

INSDOCID: <WO___9611947A1_IA>

SUBSTITUTE SHEET (RULE 26)

```
3784  ACA ATG GGG AAT CTT ACT ATT CGT GAT TTC TTG CGA ATT GGT AAT GTT CGC ATT GTT CCT  3843
1257  T   M   G   N   L   T   I   R   D   F   L   R   I   G   N   V   R   I   V   P  1276

3844  GAC CCA GTG AAT AAA ACG GTT AAA TTT GAA TGG GTT GAA TAA GAGGTATT ATG GAA AAA TTT 3905
1277  D   P   V   N   K   T   V   K   F   E   W   V   E   *           M   E   K   F   4

3906  ATG GCC GAG ATT TGG ACA AGG ATA TGT CCA AAC GCC ATT TTA TCG GAA AGT AAT TCA GTA  3965
   5  M   A   E   I   W   T   R   I   C   P   N   A   I   L   S   E   S   N   S   V  24

3966  AGA TAT AAA ATA AGT ATA GCG GGT TCT TGC CCG CTT TCT ACA GCA GGA CCA TCA TAT GTT  4025
  25  R   Y   K   I   S   I   A   G   S   C   P   L   S   T   A   G   P   S   Y   V  44

4026  AAA TTT CAG GAT AAT CCT GTA GGA AGT CAA ACA TTT AGG CGC AGG CCT TCA TTT AAG AGT  4085
  45  K   F   Q   D   N   P   V   G   S   Q   T   F   R   R   R   P   S   F   K   S  64

4086  TTT TGA CCCTTCCACCCGAGCATTAGTTGATAGTAAGTCAT ATG CTT TTT CGA CTT CAA ATG ATA CTA  4153
  65  F   *                                       M   L   F   R   L   Q   M   I   L   9

4154  CAT CAG CTG CTT TTG TTA GTT TTC ATG AAT TCT TTG ACG AAT AAT CGA ATT GTT GCT ATA  4213
  10  H   Q   L   L   L   L   V   F   M   N   S   L   T   N   N   R   I   V   A   I  29

4214  TTA ACT AGT GGA AAG GTT AAT TTT CCT CCT GAA GTA GTA TCT TGG TTA AGA ACC GCC GGA  4273
  30  L   T   S   G   K   V   N   F   P   P   E   V   V   S   W   L   R   T   A   G  49

4274  ACG TCT GCC TTT CCA TCT GAT TCT ATA TTG TCA AGA TTT GAC GTA TCA TAT GCT GCT TTT  4333
  50  T   S   A   F   P   S   D   S   I   L   S   R   F   D   V   S   Y   A   A   F  69
```

FIG.7H

<SDOCID: <WO 9611947A1_IA>

```
4334  TAT ACT TCT TCT AAA AGA GCT ATC GCA TTA GAG CAT GTT AAA CTG AGT AAT AGA AAA AGC  4393
  70  Y   T   S   S   K   R   A   I   A   L   E   H   V   K   L   S   N   R   K   S    89

4394  ACA GAT GAT TAT CAA ACT ATT TTA GAT GTT GTA TTT GAC AGT TTA GAA GAT GTA GGA GCT  4453
  90  T   D   D   Y   Q   T   I   L   D   V   V   F   D   S   L   E   D   V   G   A   109

4454  ACC GGG TTT CCA AGA AGA ACG TAT GAA AGT GTT GAG CAA TTC ATG TCG GCA GTT GGT GGA  4513
 110  T   G   F   P   R   R   T   Y   E   S   V   E   Q   F   M   S   A   V   G   G   129

4514  ACT AAT AAC GAA ATT GCG AGA TTG CCA ACT TCA GCT GCT ATA AGT AAA TTA TCT GAT TAT  4573
 130  T   N   N   E   I   A   R   L   P   T   S   A   A   I   S   K   L   S   D   Y   149

4574  AAT TTA ATT CCT GGA GAT GTT CTT TAT CTT AAA GCT CAG TTA TAT GCT GAT GCT GAT TTA  4633
 150  N   L   I   P   G   D   V   L   Y   L   K   A   Q   L   Y   A   D   A   D   L   169

4634  CTT GCT CTT GGA ACT ACA AAT ATA TCT ATC CGT TTT TAT AAT GCA TCT AAC GGA TAT ATT  4693
 170  L   A   L   G   T   T   N   I   S   I   R   F   Y   N   A   S   N   G   Y   I   189

4694  TCT TCA ACA CAA GCT GAA TTT ACT GGG CAA GCT GGG TCA TGG GAA TTA AAG GAA GAT TAT  4753
 190  S   S   T   Q   A   E   F   T   G   Q   A   G   S   W   E   L   K   E   D   Y   209

4754  GTA GTT GTT CCA GAA AAC GCA GTA GGA TTT ACG ATA TAC GCA CAG AGA ACT GCA CAA GCT  4813
 210  V   V   V   P   E   N   A   V   G   F   T   I   Y   A   Q   R   T   A   Q   A   229

4814  GGC CAA GGT GGC ATG AGA AAT TTA AGC TTT TCT GAA GTA TCA AGA AAT GGC GGC ATT TCG  4873
 230  G   Q   G   G   M   R   N   L   S   F   S   E   V   S   R   N   G   G   I   S   249
```

FIG.7I

ISDOCID: <WO___9611947A1_IA>

SUBSTITUTE SHEET (RULE 26)

```
4874  AAA CCT GCT GAA TTT GGC GTC AAT GGT ATT CGT GTT AAT TAT ATC TGC GAA TCC GCT TCA  4933
 250  K   P   A   E   F   G   V   N   G   I   R   V   N   Y   I   C   E   S   A   S   269

4934  CCT CCG GAT ATA ATG GTA CTT CCT ACG CAA GCA TCG TCT AAA ACT GGT AAA GTG TTT GGG  4993
 270  P   P   D   I   M   V   L   P   T   Q   A   S   S   K   T   G   K   V   F   G   289

4994  CAA GAA TTT AGA GAA GTT TAA ATTGAGGGACCCTTCGGGTTCCCTTTTTCTTTATAAATACTATTCAAATAAA 5066
 290  Q   E   F   R   E   V   *                                                       296

5067  GGGGCATACA ATG GCT GAT TTA AAA GTA GGT TCA ACA ACT GGA GGC TCT GTC ATT TGG CAT   5127
   1             M   A   D   L   K   V   G   S   T   T   G   G   S   V   I   W   H    17

5128  CAA GGA AAT TTT CCA TTG AAT CCA GCC GGT GAC GAT GTA CTC TAT AAA TCA TTT AAA ATA  5187
  18  Q   G   N   F   P   L   N   P   A   G   D   D   V   L   Y   K   S   F   K   I    37

5188  TAT TCA GAA TAT AAC AAA CCA CAA GCT GCT GAT AAC GAT TTC GTT TCT AAA GCT AAT GGT  5247
  38  Y   S   E   Y   N   K   P   Q   A   A   D   N   D   F   V   S   K   A   N   G    57

5248  GGT ACT TAT GCA TCA AAG GTA ACA TTT AAC GCT GGC ATT CAA GTC CCA TAT GCT CCA AAC  5307
  58  G   T   Y   A   S   K   V   T   F   N   A   G   I   Q   V   P   Y   A   P   N    77

5308  ATC ATG AGC CCA TGC GGG ATT TAT GGG GGT AAC GGT GAT GGT GCT ACT TTT GAT AAA GCA  5367
  78  I   M   S   P   C   G   I   Y   G   G   N   G   D   G   A   T   F   D   K   A    97

5368  AAT ATC GAT ATT GTT TCA TGG TAT GGC GTA GGA TTT AAA TCG TCA TTT GGT TCA ACA GGC  5427
  98  N   I   D   I   V   S   W   Y   G   V   G   F   K   S   S   F   G   S   T   G   117
```

FIG.7J

SUBSTITUTE SHEET (RULE 26)

```
5428  CGA ACT GTT GTA ATT AAT ACA CGC AAT GGT GAT ATT AAC ACA AAA GGT GTT GTG TCG GCA  5487
 118  R   T   V   V   I   N   T   R   N   G   D   I   N   T   K   G   V   V   S   A    137

5488  GCT GGT CAA GTA AGA AGT GGT GCG GCT GCT CCT ATA GCA GCG AAT GAC CTT ACT AGA AAG  5547
 138  A   G   Q   V   R   S   G   A   A   A   P   I   A   A   N   D   L   T   R   K    157

5548  GAC TAT GTT GAT GGA GCA ATA AAT ACT GTT ACT GCA AAT GCA AAC TCT AGG GTG CTA CGG  5607
 158  D   Y   V   D   G   A   I   N   T   V   T   A   N   A   N   S   R   V   L   R    177

5608  TCT GGT GAC ACC ATG ACA GGT AAT TTA ACA GCG CCA AAC TTT TTC TCG CAG AAT CCT GCA  5667
 178  S   G   D   T   M   T   G   N   L   T   A   P   N   F   F   S   Q   N   P   A    197

5668  TCT CAA CCC TCA CAC GTT CCA CGA TTT GAC CAA ATC GTA ATT AAG GAT TCT GTT CAA GAT  5727
 198  S   Q   P   S   H   V   P   R   F   D   Q   I   V   I   K   D   S   V   Q   D    217

5728  TTC GGC TAT TAT TAA GAGGACTT ATG GCT ACT TTA AAA CAA ATA CAA TTT AAA AGA AGC AAA 5789
 218  F   G   Y   Y   *          M   A   T   L   K   Q   I   Q   F   K   R   S   K    13

5790  ATC GCA GGA ACA CGT CCT GCT GCT TCA GTA TTA GCC GAA GGT GAA TTG GCT ATA AAC TTA  5849
  14  I   A   G   T   R   P   A   A   S   V   L   A   E   G   E   L   A   I   N   L    33

5850  AAA GAT AGA ACA ATT TTT ACT AAA GAT GAT TCA GGA AAT ATC ATC GAT CTA GGT TTT GCT  5909
  34  K   D   R   T   I   F   T   K   D   D   S   G   N   I   I   D   L   G   F   A    81

5910  AAA GGC GGG CAA GTT GAT GGC AAC GTT ACT ATT AAC GGA CTT TTG AGA TTA AAT GGC GAT  5969
  54  K   G   G   Q   V   D   G   N   V   T   I   N   G   L   L   R   L   N   G   D    73
```

FIG.7K

SUBSTITUTE SHEET (RULE 26)

SDOCID: <WO___9611947A1_IA>

```
5970  TAT GTA CAA ACA GGT GGA ATG ACT GTA AAC GGA CCC ATT GGT TCT ACT GAT GGC GTC ACT  6029
  74  Y   V   Q   T   G   G   M   T   V   N   G   P   I   G   S   T   D   G   V   T    93

6030  GGA AAA ATT TTC AGA TCT ACA CAG GGT TCA TTT TAT GCA AGA GCA ACA AAC GAT ACT TCA  6089
  94  G   K   I   F   R   S   T   Q   G   S   F   Y   A   R   A   T   N   D   T   S   113

6090  AAT GCC CAT TTA TGG TTT GAA AAT GCC GAT GGC ACT GAA CGT GGC GTT ATA TAT GCT CGC  6149
 114  N   A   H   L   W   F   E   N   A   D   G   T   E   R   G   V   I   Y   A   R   133

6150  CCT CAA ACT ACA ACT GAC GGT GAA ATA CGC CTT AGG GTT AGA CAA GGA ACA GGA AGC ACT  6209
 134  P   Q   T   T   T   D   G   E   I   R   L   R   V   R   Q   G   T   G   S   T   153

6210  GCC AAC AGT GAA TTC TAT TTC CGC TCT ATA AAT GGA GGC GAA TTT CAG GCT AAC CGT ATT  6269
 154  A   N   S   E   F   Y   F   R   S   I   N   G   G   E   F   Q   A   N   R   I   173

6270  TTA GCA TCA GAT TCG TTA GTA ACA AAA CGC ATT GCG GTT GAT ACC GTT ATT CAT GAT GCC  6329
 174  L   A   S   D   S   L   V   T   K   R   I   A   V   D   T   V   I   H   D   A   193

6330  AAA GCA TTT GGA CAA TAT GAT TCT CAC TCT TTG GTT AAT TAT GTT TAT CCT GGA ACC GGT  6389
 194  K   A   F   G   Q   Y   D   S   H   S   L   V   N   Y   V   Y   P   G   T   G   213

6390  GAA ACA AAT GGT GTA AAC TAT CTT CGT AAA GTT CGC GCT AAG TCC GGT GGT ACA ATT TAT  6449
 214  E   T   N   G   V   N   Y   L   R   K   V   R   A   K   S   G   G   T   I   Y   911

6450  CAT GAA ATT GTT ACT GCA CAA ACA GGC CTG GCT GAT GAA GTT TCT TGG TGG TCT GGT GAT  6509
 234  H   E   I   V   T   A   Q   T   G   L   A   D   E   V   S   W   W   S   G   D   253
```

FIG.7L

SUBSTITUTE SHEET (RULE 26)

\BDOCID: <WO___9611947A1_IA>

```
6510  ACA CCA GTA TTT AAA CTA TAC GGT ATT CGT GAC GAT GGC AGA ATG ATT ATC CGT AAT AGC  6569
 254  T   P   V   F   K   L   Y   G   I   R   D   D   G   R   M   I   I   R   N   S    273

6570  CTT GCA TTA GGT ACA TTC ACT ACA AAT TTC CCG TCT.AGT GAT TAT GGC AAC GTC GGT GTA  6629
 274  L   A   L   G   T   F   T   T   N   F   P   S   S   D   Y   G   N   V   G   V    293

6630  ATG GGC GAT AAG TAT CTT GTT CTC GGC GAC ACT GTA ACT GGC TTG TCA TAC AAA AAA ACT  6689
 294  M   G   D   K   Y   L   V   L   G   D   T   V   T   G   L   S   Y   K   K   T    313

6690  GGT GTA TTT GAT CTA GTT GGC GGT GGA TAT TCT GTT GCT TCT ATT ACT CCT GAC AGT TTC  6749
 314  G   V   F   D   L   V   G   G   G   Y   S   V   A   S   I   T   P   D   S   F    333

6750  CGT AGT ACT CGT AAA GGT ATA TTT GGT CGT TCT GAG GAC CAA GGC GCA ACT TGG ATA ATG  6809
 334  R   S   T   R   K   G   I   F   G   R   S   E   D   Q   G   A   T   W   I   M    353

6810  CCT GGT ACA AAT GCT GCT CTC TTG TCT GTT CAA ACA CAA GCT GAT AAT AAC AAT GCT GGA  6869
 354  P   G   T   N   A   A   L   L   S   V   Q   T   Q   A   D   N   N   N   A   G    373

6870  GAC GGA CAA ACC CAT ATC GGG TAC AAT GCT GGC GGT AAA ATG AAC CAC TAT TTC CGT GGT  6929
 374  D   G   Q   T   H   I   G   Y   N   A   G   G   K   M   N   H   Y   F   R   G    393

6930  ACA GGT CAG ATG AAT ATC AAT ACC CAA CAA GGT ATG GAA ATT AAC CCG GGT ATT TTG AAA  6989
 394  T   G   Q   M   N   I   N   T   Q   Q   G   M   E   I   N   P   G   I   L   K    413

6990  TTG GTA ACT GGC TCT AAT AAT GTA CAA TTT TAC GCT GAC GGA ACT ATT TCT TCC ATT CAA  7049
 414  L   V   T   G   S   N   N   V   Q   F   Y   A   D   G   T   I   S   S   I   Q    433
```

FIG.7M

WSDOCID: <WO___9611947A1_IA>

SUBSTITUTE SHEET (RULE 26)

```
7050  CCT ATT AAA TTA GAT AAC GAG ATA TTT TTA ACT AAA TCT AAT AAT ACT GCG GGT CTT AAA  7109
 434  P   I   K   L   D   N   E   I   F   L   T   K   S   N   N   T   A   G   L   K    453

7110  TTT GGA GCT CCT AGC CAA GTT GAT GGC ACA AGG ACT ATC CAA TGG AAC GGT GGT ACT CGC  7169
 454  F   G   A   P   S   Q   V   D   G   T   R   T   I   Q   W   N   G   G   T   R    473

7170  GAA GGA CAG AAT AAA AAC TAT GTG ATT ATT AAA GCA TGG GGT AAC TCA TTT AAT GCC ACT  7229
 474  E   G   Q   N   K   N   Y   V   I   I   K   A   W   G   N   S   F   N   A   T    493

7230  GGT GAT AGA TCT CGC GAA ACG GTT TTC CAA GTA TCA GAT AGT CAA GGA TAT TAT TTT TAT  7289
 494  G   D   R   S   R   E   T   V   F   Q   V   S   D   S   Q   G   Y   Y   F   Y    513

7290  GCT CAT CGT AAA GCT CCA ACC GGC GAC GAA ACT ATT GGA CGT ATT GAA GCT CAA TTT GCT  7349
 514  A   H   R   K   A   P   T   G   D   E   T   I   G   R   I   E   A   Q   F   A    533

7350  GGG GAT GTT TAT GCT AAA GGT ATT ATT GCC AAC GGA AAT TTT AGA GTT GTT GGG TCA AGC  7409
 534  G   D   V   Y   A   K   G   I   I   A   N   G   N   F   R   V   V   G   S   S    553

7410  GCT TTA GCC GGC AAT GTT ACT ATG TCT AAC GGT TTG TTT GTC CAA GGT GGT TCT TCT ATT  7469
 554  A   L   A   G   N   V   T   M   S   N   G   L   F   V   Q   G   G   S   S   I    573

7470  ACT GGA CAA GTT AAA ATT GGC GGA ACA GCA AAC GCA CTG AGA ATT TGG AAC GCT GAA TAT  7529
 574  T   G   Q   V   K   I   G   G   T   A   N   A   L   R   I   W   N   A   E   Y    593

7530  GGT GCT ATT TTC CGT CGT TCG GAA AGT AAC TTT TAT ATT ATT CCA ACC AAT CAA AAT GAA  7589
 594  G   A   I   F   R   R   S   E   S   N   F   Y   I   I   P   T   N   Q   N   E    613
```

FIG.7N

SUBSTITUTE SHEET (RULE 26)

RESDOCID: <WO 9611947A1_IA>

```
7590  GGA GAA AGT GGA GAC ATT CAC AGC TCT TTG AGA CCT GTG AGA ATA GGA TTA AAC GAT GGC  7649
 614  G   E   S   G   D   I   H   S   S   L   R   P   V   R   I   G   L   N   D   G   633

7650  ATG GTT GGG TTA GGA AGA GAT TCT TTT ATA GTA GAT CAA AAT AAT GCT TTA ACT ACG ATA  7709
 634  M   V   G   L   G   R   D   S   F   I   V   D   Q   N   N   A   L   T   T   I   653

7710  AAC AGT AAC TCT CGC ATT AAT GCC AAC TTT AGA ATG CAA TTG GGG CAG TCG GCA TAC ATT  7769
 654  N   S   N   S   R   I   N   A   N   F   R   M   Q   L   G   Q   S   A   Y   I   673

7770  GAT GCA GAA TGT ACT GAT GCT GTT CGC CCG GCG GGT GCA GGT TCA TTT GCT TCC CAG AAT  7829
 674  D   A   E   C   T   D   A   V   R   P   A   G   A   G   S   F   A   S   Q   N   693

7830  AAT GAA GAC GTC CGT GCG CCG TTC TAT ATG AAT ATT GAT AGA ACT GAT GCT AGT GCA TAT  7889
 694  N   E   D   V   R   A   P   F   Y   M   N   I   D   R   T   D   A   S   A   Y   713

7890  GTT CCT ATT TTG AAA CAA CGT TAT GTT CAA GGC AAT GGC TGC TAT TCA TTA GGG ACT TTA  7949
 714  V   P   I   L   K   Q   R   Y   V   Q   G   N   G   C   Y   S   L   G   T   L   733

7950  ATT AAT AAT GGT AAT TTC CGA GTT CAT TAC CAT GGC GGC GGA GAT AAC GGT TCT ACA GGT  8009
 734  I   N   N   G   N   F   R   V   H   Y   H   G   G   G   D   N   G   S   T   G   753

8010  CCA CAG ACT GCT GAT TTT GGA TGG GAA TTT ATT AAA AAC GGT GAT TTT ATT TCA CCT CGC  8069
 754  P   Q   T   A   D   F   G   W   E   F   I   K   N   G   D   F   I   S   P   R   773

8070  GAT TTA ATA GCA GGC AAA GTC AGA TTT GAT AGA ACT GGT AAT ATC ACT GGT GGT TCT GGT  8129
 774  D   L   I   A   G   K   V   R   F   D   R   T   G   N   I   T   G   G   S   G   793
```

FIG.70

BNSDOCID: <WO___9611947A1_IA>

SUBSTITUTE SHEET (RULE 26)

```
8130  AAT TTT GCT AAC TTA AAC AGT ACA ATT GAA TCA CTT AAA ACT GAT ATC ATG TCG AGT TAC  8180
 794  N   F   A   N   L   N   S   T   I   E   S   L   K   T   D   I   M   S   S   Y    813

8190  CCA ATT GGT GCT CCG ATT CCT TGG CCG AGT GAT TCA GTT CCT GCT GGA TTT GCT TTG ATG  8249
 814  P   I   G   A   P   I   P   W   P   S   D   S   V   P   A   G   F   A   L   M    833

8250  GAA GGT CAG ACC TTT GAT AAG TCC GCA TAT CCA AAG TTA GCT GTT GCA TAT CCT AGC GGT  8309
 834  E   G   Q   T   F   D   K   S   A   Y   P   K   L   A   V   A   Y   P   S   G    853

8310  GTT ATT CCA GAT ATG CGC GGG CAA ACT ATC AAG GGT AAA CCA AGT GGT CGT GCT GTT TTG  8369
 854  V   I   P   D   M   R   G   Q   T   I   K   G   K   P   S   G   R   A   V   L    873

8370  AGC GCT GAG GCA GAT GGT GTT AAG GCT CAT AGC CAT AGT GCA TCG GCT TCA AGT ACT GAC  8429
 874  S   A   E   A   D   G   V   K   A   H   S   H   S   A   S   A   S   S   T   D    893

8430  TTA GGT ACT AAA ACC ACA TCA AGC TTT GAC TAT GGT ACG AAG GGA ACT AAC AGT ACG GGT  8489
 894  L   G   T   K   T   T   S   S   F   D   Y   G   T   K   G   T   N   S   T   G    913

8490  GGA CAC ACT CAC TCT GGT AGT GGT TCT ACT AGC ACA AAT GGT GAG CAC AGC CAC TAC ATC  8549
 914  G   H   T   H   S   G   S   G   S   T   S   T   N   G   E   H   S   H   Y   I    933

8550  GAG GCA TGG AAT GGT ACT GGT GTA GGT GGT AAT AAG ATG TCA TCA TAT GCC ATA TCA TAC  8609
 934  E   A   W   N   G   T   G   V   G   G   N   K   M   S   S   Y   A   I   S   Y    953

8610  AGG GCG GGT GGG AGT AAC ACT AAT GCA GCA GGG AAC CAC AGT CAC ACT TTC TCT TTT GGG  8669
 954  R   A   G   G   S   N   T   N   A   A   G   N   H   S   H   T   F   S   F   G    973
```

FIG.7P

NSDOCID: <WO 9611947A1 IA>

```
8670  ACT AGC AGT GCT GGC GAC CAT TCC CAC TCT GTA GGT ATT GGT GCT CAT ACC CAC ACG GTA  8729
 974  T   S   S   A   G   D   H   S   H   S   V   G   I   G   A   H   T   H   T   V    993

8730  GCA ATT GGA TCA CAT GGT CAT ACT ATC ACT GTA AAT AGT ACA GGT AAT ACA GAA AAC ACG  8789
 994  A   I   G   S   H   G   H   T   I   T   V   N   S   T   G   N   T   E   N   T   1013

8790  GTT AAA AAC ATT GCT TTT AAC TAT ATC GTT CGT TTA GCA TAA GGAGAGGGGCTTCGGCCCTTCTAA  8855
1014  V   K   N   I   A   F   N   Y   I   V   R   L   A   •                          1027
```

FIG.7Q

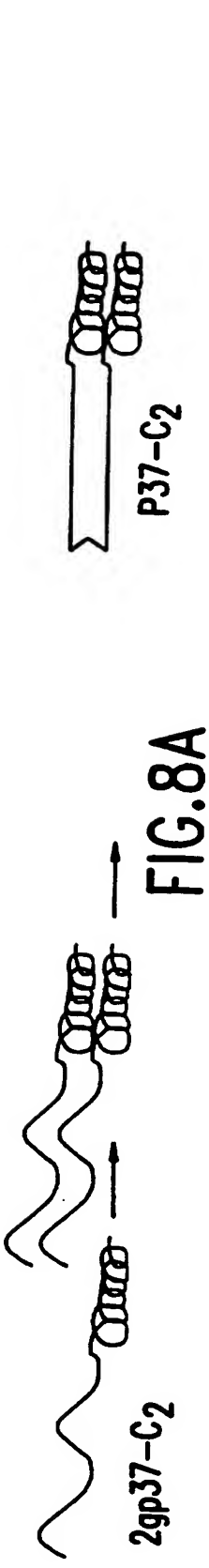


FIG.8A

FIG.8B

FIG.9

WSDOCID: <WO___9611947A1_IA>

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/13023

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07K 14/195; C12P 21/06; C07H 17/00
US CL :530/300, 350; 435/69.1, 69.7; 536/23.1, 23.4, 23.7
According to International Patent Classification (IPC) r to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/300, 350; 435/69.1, 69.7; 536/23.1, 23.4, 23.7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Dialog

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Science, Vol. 254, issued 29 November 1991, D.H. Freedman, "Exploiting the nanotechnology of life", pages 1308-1310, see entire document.	1-53
A	Science, Vol. 254, issued 29 November 1991, G.M. Whitesides et al., "Molecular self-assembly and nanochemistry: A chemical strategy for the synthesis of nanostructures", pages 1312-1319, see entire document.	1-53
A	Genetics, Vol. 94, issued March 1980, J.N. Levy et al., "Region-specific recombination in phage T4. II. Structure of the recombinants", pages 531-547, see entire document.	1-53

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date f the actual completion of the international search
16 JANUARY 1996

Date of mailing of th international search report
01 FEB 1996

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Auth rized officer
Karen Cochrane Carlson, Ph.D.
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

SDOCID: <WO___9611947A1_IA>

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	J. Mol. Biol., Vol. 132, issued 1979, W.C. Earnshaw et al., "The distal half of the tail fibre of bacteriophage T4 rigidly linked domains and cross-β structure", pages 101-131, see entire document.	1-53

SDOCID: <WO___9611947A1_IA>

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 March 2001 (29.03.2001)

PCT

(10) International Publication Number
WO 01/21646 A1

(51) International Patent Classification[7]: C07K 14/00, 1/113, G11B 9/00

(21) International Application Number: PCT/GB00/03576

(22) International Filing Date:
18 September 2000 (18.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9922013.9 17 September 1999 (17.09.1999) GB

(71) Applicant *(for all designated States except US)*: UNIVERSITY OF SUSSEX [GB/GB]; Falmer, Brighton, Sussex BN1 9QG (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: WOOLFSON, Derek, N. [GB/GB]; University of Sussex, Falmer, Brighton, Sussex BN1 9QG (GB). WALSHAW, John [GB/GB]; University of Sussex, Falmer, Brighton, Sussex BN1 9QG (GB). PANDYA, Maya, J. [GB/GB]; University of Sussex, Falmer, Brighton, Sussex BN1 9QG (GB). COLYER, John [GB/GB]; Elfordlea, Mill Lane, Bardsey LS17 9AN (GB).

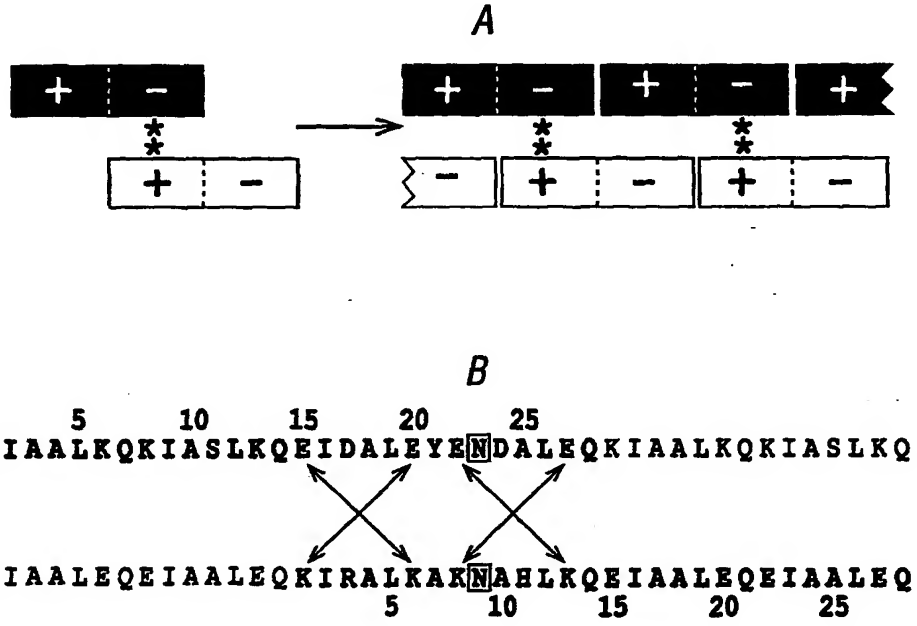
(74) Agents: DEAN, John, Paul et al.; Withers & Rogers, Goldings House, 2 Hays Lane, London SE1 2HW (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: PROTEIN STRUCTURES AND PROTEIN FIBRES



(57) Abstract: This invention relates to protein fibre formation and in particular to methods of producing protein fibres to form a protein structure comprising a plurality of first polypeptide units arranged in a first polypeptide strand and a plurality of second polypeptide units arranged in a second polypeptide strand, the strands preferably forming a coiled coil structure, and in which a first polypeptide unit in the first strand extends beyond a corresponding second polypeptide unit in the second strand in the direction of the strands.

WO 01/21646 A1

Published:
— *With international search report.*
— *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

PROTEIN STRUCTURES AND PROTEIN FIBRES

This invention relates to protein structures, to methods of producing those protein structures, and to protein fibres and other materials and assemblies produced using those protein structures.

The process of molecular self-assembly is central to all biological systems and is assuming increasing importance and application in biotechnology (L. Q. Gu, *et al* (1999) *Nature* **398**, 686) and nanotechnology (K. E. Drexler, (1999) TIBTECH **17**, 5). The characterization of natural biomolecular assemblies motivates and directs the development of model self-assembling systems and, in turn, these advance our understanding of biology. For proteins at least, the coiled coil is arguably the simplest self-assembling system. Coiled coils are protein-folding motifs that direct and cement a wide variety of protein-protein interactions (A. Lupas, (1996) *Trends Biochem. Sci* **21**, 375). In structural terms, coiled coils are relatively straightforward: they are α-helical bundles with between 2 and 5 strands that can be arranged in parallel, antiparallel or mixed topologies. The basic sequence features that guide the formation of coiled coils from peptides are reasonably well understood (P. B. Harbury *et al* (1993) *Science* **262**, 1401; D. N. Woolfson and T. Alber (1995) *Protein Sci.* **4**, 1596). For instance, most coiled-coil sequences are dominated by a 7-residue repeat of hydrophobic (H) and polar (P) residues, $(HPPHPPP)_n$, known as the "heptad repeat". When configured into an α-helix this pattern gives an amphipathic structure, the hydrophobic face of which directs oligomer-assembly. Furthermore, both the number and the direction of chains within a coiled-coil bundle is determined predominantly by residues that form or flank the hydrophobic core namely, residues at the first, fourth, fifth and seventh positions of the heptad repeat. For instance, coiled coils which form dimers (i.e. two-stranded assemblies) usually have isoleucine or valine residues at the first position and a leucine residue at the fourth position. By contrast, coiled coils that form trimers (i.e. three-stranded assemblies) often have the same residues (i.e both isoleucine or both leucine) at both "H" positions. Finally, hetero-oligomers (that is coiled coils made from strands with different amino-acid sequences) may be directed by complementary charged interactions that flank the hydrophobic core. For these reasons, there have been a number of successful *de novo* protein designs based on the coiled coil These include some ambitious structures that extend the natural repertoire of coiled-coil motifs (S. Nautiyal *et al* (1995) *Biochemistry* **34**, 11645; A. Lombardi *et al* (1996)

Biopolymers **40**, 495; D. H. Lee *et al* (1996) *Nature* **382**, 525; P. B. Harbury *et al* (1998) *Science* **282**, 1462; J. P. Schneider *et al* (1998) *Folding Des.* **3**, R29).

In addition to commonly accepted structures with a single, contiguous heptad repeat, the inventors have identified sequences with multiple, offset heptad repeats which help explain oligomer-state specification in coiled coils. For example, sequences with two heptad repeats offset by two residues; i.e *a/f-b/g'-c/a'-d/b'-e/c'-f/d'-g/e'* set up two hydrophobic seams on opposite sides of the helix formed. Such helices may combine to bury these hydrophobic surfaces in two different ways and form two distinct structures: open "α-sheets" and closed "α-cylinders".

Other relevant aspects of coiled-coil structure are described in WO99/11774, the disclosure of which is incorporated herein by way of reference.

This understanding of coiled coils, and the resulting protein designs, centres on short structures as exemplified by the leucine-zipper motifs (E. K. O'Shea *et al* (1989) *Science* **243**, 538; E. K. O'Shea *et al* (1991) *Science* **254**, 539), which are found in a variety of transcription factors. In contrast, most natural coiled coils extend over hundreds of amino acids (A. Lupas (1996) *supra*; J. Sodek *et al* (1972) *Proc. Natl. Acad. Sci. U.S.A* **69**, 3800) and many assemble further to form thicker, multi-stranded filaments (H. Herrmann and U. Aebi (1998) *Curr. Opin. Struct. Biol.* **8**, 177).

With the goal of making elongated structures to improve our understanding of coiled coils, and to develop protein-design studies, we initially designed two 28-residue peptides — dubbed Self-Assembling Fibre peptides, SAF-p1 and SAF-p2 — to fold and form extended fibres when mixed. Focusing on the buried, hydrophobic-core positions of the structure, rules were incorporated to direct parallel dimer formation and to guard against alternative oligomers and topologies (P. B. Harbury *et al* (1993) *supra*; D. N. Woolfson and T. Alber (1995) *supra*; L. J. Gonzalez *et al* (1996) *Nature Struct. Biol.* **3**, 1011). The building block of the design was a staggered heterodimer with overhanging or "sticky" ends. This contrasts with and distinguishes it from the natural and designer coiled-coil assemblies that have been characterized to date, in which the polypeptide strands align in-register, i.e they have blunt or "flush" ends. Complementary core interactions and flanking ion-pairs were incorporated into the overhangs to facilitate longitudinal association of the heterodimers (Figs. 1&2). This

principle of using "sticky ends" is well developed in molecular biology for assembling DNA (S. J. Palmer *et al* (1998) *Nucleic Acids Res.* **26**, 2560), and has been used to design intricate DNA crystals (E. Winfree *et al* (1998) *Nature* **394**, 539). However, to our knowledge, our application of sticky end-directed molecular assembly to peptides is new; although we do note that head-to-tail packing of helices has been observed in recently solved crystal structures for two designer peptides (N. L. Ogihara *et al* (1997) *Protein Sci.* **6**, 80; G. G. Prive *et al* (1999) *Protein Sci.* **8**, 1400). These were helical peptides that crystallised with their helical ends in contact so as to form pseudo-continuous helices in the solid state. In other words they formed "blunt-ended" arrangements.

According to one aspect of the invention there is provided a protein structure comprising a plurality of first peptide monomer units arranged in a first strand and a plurality of second peptide monomer units arranged in a second strand, the strands preferably forming a coiled-coil structure, and in which a first peptide monomer unit in the first strand extends beyond a corresponding second peptide monomer unit in the second strand in the direction of the strands. The protein structures of the invention have numerous advantages. For example, relatively long protein fibres can be formed with little material - 1 μl of a 100 μM solution of the peptide monomers may provide enough material to form 10 m of fibre 50 nm thick.

At least one charged amino acid residue of the first peptide monomer unit may be arranged to attract an oppositely-charged amino acid residue of the second peptide monomer unit. Preferably, the charged amino acid residue is in an end portion of the first peptide monomer unit, which extends beyond the corresponding second peptide monomer unit in the second strand. At least one strand may consist solely of first or second peptide monomer units respectively i.e homogenous strands. Heterologous strands are also contemplated. The peptide monomer units may comprise a repeating structural unit. Preferably, the repeating structural unit comprises a heptad repeat motif, having the pattern:

```
hpphppp
abcdefg
```

Preferably, the repeat may include isoleucine or asparagine at position a and leucine at position d. Other repeats (e.g hendecads - abcdefghijk) and amino acid compositions may also be used (see WO99/11774).

Preferably, the heptad repeat comprises oppositely-charged residues at positions e and g respectively. The oppositely-charged residues may be, for example, glutamic acid and lysine residues or arginine and aspartic acid. The use of synthetic amino acids, such as ornithine is also envisaged.

A protein structure in accordance with the invention may be also specified by pairs of asparagine residues in the "a" positions provided by corresponding first and second peptide monomer units.

In a preferred protein structure, the first and second peptide monomer units have the following sequences:

a) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1C) and

b) KIRALKAKNAHLKQEIAALEQEIAALEQ (SAF-p2D) respectively; or

c) KIAALKQKIAALKQEIDALEYENDALEQ (SAF-p1A) and

d) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C) respectively; or

e) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1C) and

f) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C) respectively.

It will be appreciated that these are examples only of 4-heptad structures and that other lengths are possible and envisaged for use in the invention.

According to another aspect of the invention, there is provided a method of producing protein structures, the method comprising providing a mixture of first and second peptide monomer units which associate to form a protein structure according to the invention. The structure can be derivatised and/or stabilized by cross-linking.

Derivatization of the peptide monomer units before or after assembly into the protein structures of the invention may be performed. For example, fluorescent moieties (fluorophores) may be attached to the coiled coil as described in WO99/11774. The addition of fluorescent moieties may assist visualization of the protein structure. Substitution with

functional groups at the "f" position in the heptad repeat is especially preferred as that position is on the outside of the helix (see Fig. 1C and 1E). Other derivatives may include attaching binders to the peptide monomer units for example so that units which can bind other entities can be produced.

The first and second peptide monomers and the strands may have the characteristics described above.

The invention also provides protein fibres produced by an association of protein structures according to the invention.

The protein structures may also be arranged to form tubular structures. In particular, the structures may be arranged to form nanotubes.

According to another aspect of the invention, there is provided a kit for making protein structures, the kit comprising first and second peptide monomer units which associate to form a protein structure or protein fibres according to the invention.

The protein structures of the invention may be assembled in two and three dimensional arrays. For example, two dimensional mats can be formed which can function, for example as filters. Three dimensional grids or matrices can also be formed again, for example, for use as sieves or filters or for organising other associated or conjugated molecules in three dimensions.

In a preferred embodiment, a matrix is assembled *in situ*. For example, a matrix can be formed in a solution to entrap contaminants in the solution and then the matrix, together with contaminants, can be removed from the solution for example by centrifugation.

The stability of the protein structures at higher temperatures may be improved by making the peptide monomers longer, such that the overlap between corresponding first and second monomer unit residues is increased. Increases in monomer length have previously been shown to stabilize coiled coil structures. Alternatively, stability can be improved by introducing bonding between adjacent peptide monomer units in the same strand. For example, Kent (Dawson *et al* (1994) Science **266**: 776) and co-workers have produced peptide bonds between adjacent polypeptide units by coupling and subsequent rearrangement of a

cysteine residue at the N end of one polypeptide unit to a thio-ester derivatised C-terminus of another unit.

Additionally, the protein structures may be stabilised and derivatised by using them to template the polymerisation of synthetic polymers.

Definitions

The terms used in the specification are to be given the ordinary meaning attributed to them by the skilled addressee. The following is given by way of clarification:

Amino acid.

This term embraces both naturally-occuring amino acids and synthetic amino acids as well as naturally-occuring amino acids which have been modified in some way to alter certain properties such as charge. In all cases references to naturally-occurring amino acids may be considered to include synthetic amino acids which may be substituted therefor.

Coiled Coil

A coiled-coil is a peptide/protein sequence usually with a contiguous pattern of hydrophobic residues spaced 3 and 4 residues apart, which assembles (folds) to form a multi-meric bundle of helices. Coiled-coils including sequences with multiple offset repeats are also contemplated.

Dimer

A dimer is a two stranded structure.

Heterodimer

A heterodimer is a dimeric structure formed by two different stands.

Staggered heterodimer

A staggered heterodimer is a structure in which the two strands assemble to leave overlapping ends that are not interacting within the heterodimer.

Blunt-end assembly

Blunt-end assembly is association where the two strands combine to give flushed i.e non-overlapping ends.

Protofibril

A protofibril is a protein structure assembled longitudinally from staggered heterodimers interacting through their overhanging ends.

Fibre

A fibre is a structure formed by lateral association of two or more protofibrils.

Protein structures and methods of producing protein structures in accordance with the invention will now be described, by way of example only, with reference to the accompanying Figures 1 to 8 in which:

Fig. 1 illustrates the design and the sequences of self-assembling fibre (SAF) peptide monomers of the invention.

Fig. 2 illustrates computer modelling of the designed self-assembling fibre of the invention.

Fig. 3 illustrates the results of circular dichroism (CD) and linear dichroism (LD) experiments on protein structures of the invention.

Fig. 4 illustrates the assembly of synthetic protein fibres visualized directly by transmission electron microscopy and an analysis of fibre width In all panels, the white scale bars represent 100 nm. Fig. 4D is a histogram showing the distribution of fibre widths determined using TEM for fresh (white bars) and matured (black bars) mixtures of SAF peptides at 100 μM (a width value of "*x*" on the histogram includes all measurements made from "(*x*–*5*) to *x*").

Fig. 5 is a cartoon showing the possible anti-typic association of parallel helical peptides leading to a homo-oligomeric peptide nanotube.

Fig. 6 is an x-ray diffraction pattern of an aligned protein fibre of the invention.

Fig. 7 is an electromicrograph showing fibres which have been derivatised through the inclusion of flurophores; and

Fig. 8 shows amino acid sequences designed to form blunt-ended heterodimers.

1) Peptide Design and Synthesis

Various peptide monomer units were designed as described above. The monomers and capping peptides (designed to complement the sticky ends of the monomers so as to produce flush, or blunt ends and, so, arrest longitudinal fibre assembly) are set out in Table 1:

SUBSTITUTE SHEET (RULE 26)

TABLE 1

PEPTIDE	SEQUENCE g abcdefg abcdefg abcdefg abcdef 5 10 15 20 25	DESIGN	CD DATA @ 10 μM	 @ 100 μM	LD DATA @ 10 μM	 @ 100 μM	EM DATA @ 10 μM	 @ 100 μM
CAP-p1A	*YGPGE IAALEQE NAALEQ	prototype	unfolded					
SAF-p1A	K IAALKQK IAALKQE IDALEYE NDALEQ	prototype; slowly precipitates	unfolded	~ 45 % α-helix				no fibres
SAF-p1B	*K IAALKQK IAALKQE IDALEYE NDALEQ*	chemical capping of the ends (↑ stability)	~ 60 % α-helix	~ 70 % α-helix				
SAF-p1C	K IAALKQK IASLKQE IDALEYE NDALEQ	no capping (↓ stability); mutate $A_{11} \rightarrow S$ (↑ solubility & ↓ helix stability)	unfolded	~ 20 % α-helix		no signal		no fibres
CAP-p2A	K IAALKQK NAALKQG GW*	prototype	unfolded					
SAF-p2A	K ISALKWK NASLKQE IAALEQE IAALEQ	prototype; low solubility	unfolded					
SAF-p2B	*K IRALKWK NAHLKQE IAALEQE IAALEQ*	mutate $S_3 \rightarrow R$ & $S_{11} \rightarrow H$ (↑ solubility & ↑ helix stability)	~ 60 % α-helix	~ 95 % α-helix				
SAF-p2C	K IRALKWK NAHLKQE IAALEQE IAALEQ	no capping (↓ stability)	unfolded	~ 20 % α-helix				thin fibres
SAF-p2D	K IRALKAK NAHLKQE IAALEQE IAALEQ	mutate $W_7 \rightarrow A$ (investigate role of Trp in fibrillogenesis)	~ 15 % α-helix	~ 45 % α-helix		no signal		no fibres
SAF-p2E	K IRALKCK NAHLKQE IAALEQE IAALEQ	mutate $A_7 \rightarrow C$ (for derivatization & cross-linking)						

*=Chemical capping = CH_3CO at the N terminus and NH_2 at the C terminus

Fig. 1 shows (A) A mechanism for self-assembly: complementary charges in "companion" peptides direct the formation of staggered, parallel heterodimers; the resulting "sticky" ends are also complementary and promote longitudinal association into extended structures. Fig 1(B) shows the designed amino acid sequences: each peptide comprised canonical heptad repeats (***abcdefg***) with Ile at ***a*** and Leu at ***d*** to guide the formation of coiled-coil dimers; oppositely-charged residues were incorporated at ***e*** and ***g*** to favour the staggered dimer with sticky ends; asparagine residues (which preferentially pairs with each other at ***a*** sites (Gonzalez L *et al* (1996) Nature Structural Biology 3, 13: 1011-1018) were included to cement the prescribed register further and to favour the parallel structures. Fig. 1(C) is a helical-wheel representation, summarizing the designed sequences in context. The view is from the N-terminus with heptad sites labeled ***a-g*** and assumes 3.5 residues per helical turn to emphasise the heptad repeat.

The peptides were synthesized on an Applied Biosystems 432A Peptide Synthesizer using solid-phase methods and Fmoc chemistry. Peptide samples were purified using reversed-phase HPLC and their identities confirmed by MALDI-TOF mass spectrometry.

Various combinations of peptide monomers and capping peptides were tested as set out in Table 2:

TABLE 2

MISC.	PEPTIDE MIXTURE	EQUILIBRATION	CD DATA @ 10 μM	CD DATA @ 100 μM	LD DATA @ 10 μM	LD DATA @ 100 m μM	EM DATA @ 10 μM	EM DATA @ 100 μM
	SAF-p1A, SAF-p2A		~ 20 % α-helix	~ 40 % α-helix $T_m \approx 30$ °C				
	SAF-p1A, SAF-p2A, CAP-p1A, CAP-p2A			~ 20 % α-helix				
	SAF-p1A, SAF-p2B	rapid		~ 65 % α-helix $T_m \approx 36/46$ °C (expect 38 °C if no interaction)				
	SAF-p1B, SAF-p2B	rapid	~ 70 % α-helix $T_m \approx 25$ °C (expect 19 °C if no interaction)	~ 80 % α-helix $T_m \approx 35$ °C (expect 36 °C if no interaction)				no fibres
	SAF-p1A, SAF-p2C	slow @ 100 μM; clouding occurs	~ 20 % α-helix	~ 80 % α-helix biphasic thermal melt			no fibres	thick fibres (~ 45 nm wide)
no signal from fibres in 1D-NMR	SAF-p1C, SAF-p2C	slow @ 100 μM; clouding occurs	~ 25 % α-helix	~ 65 % α-helix (~25% in 0.5 M salt); unusual spectral shape; no aromatic signal	no signal	strong signal from backbone & aromatics	no fibres	thick fibres (~ 45 nm wide); no fibres in 0.5 M salt
	SAF-p1C, SAF-p2D	slow @ 100 μM; clouding occurs	~ 25 % α-helix	~ 85 % α-helix (~20% in 0.5 M salt); unusual spectral shape	no signal	strong signal from backbone & aromatics; no signal in 0.5 M salt	no fibres	thick fibres (~ 45 nm wide); no fibres in 0.5 M salt

Indicates large structures are formed.

SUBSTITUTE SHEET (RULE 26)

In addition and as a control, the SAF-p1c sequence was permuted (N- and C-terminal halves were swapped) to produce peptide SAF-p3:

E IDALEYE NDALEQK IAALKQK IASLKQ

This design should combine with SAF-p2D to form a blunt-ended structure, which should not form fibres.

2) Modeling of Protein Fibre Structure

A model of the three-dimensional structure of the designed protein fibre resulting from the assembly of SAF-p1 and SAF-p2 was made from the minimised structure of a model coiled-coil 35-mer, $(LAALAAA)_5$, which was generated using Crick's Equation and had an ideally packed interface (G. Offer and R. Sessions, *J. Mol. Biol.* **249**, 967 (1995)). Copies of the 35-mer were superimposed with an overlap of one heptad repeat to extend the structural template, and the backbone was rejoined after removal of overlapping segments. Residues in the two-stranded template were replaced with the sequences of the SAF peptides, staggered relative to each other by two heptad repeats according to the alignment in Fig. 1B. The structure was soaked in a 5 Å layer of water and energy minimised until the average absolute derivative of coordinates with respect to energy fell below 0.01 kcal $Å^{-1}$. The structure was built and visualized using Insight II 97.0 (Molecular Simulations Inc.), and was energy-minimized using Discover 2.9.8 (Molecular Simulations Inc.) with the consistent valence forcefield. In Fig 2(A) peptides SAF-p1 and SAF-p2 (each coloured dark grey-to-light grey from the N-terminus) interact through core residues including asparagine pairs (coloured mid-grey) to form the two strands of a staggered, parallel, coiled-coil fibre. In Fig. 2(B), negatively charged glutamate side chains (coloured light grey) and positively charged lysine side chains (coloured black) form complementary charge interactions between the SAF peptides.

3) Circular Dichroism Experiments

Peptide samples were incubated at 5°C in 10 mM MOPS (3-(N-Morpholino)propanesulfonic acid), pH 7. Sample concentrations were determined from their UV absorbance at 280 nm (SAF-p1) and 214 nm (SAF-p2). After baseline correction, ellipticities in mdeg were converted to molar ellipticities (deg cm^2 dmol-res^{-1}) by normalizing for the concentration of

peptide bonds. Data were recorded in a cell of 1 mm path length by integrating the signal for 5s (and 1s for the fresh 100 μM peptide mixture) every nm in the range 205-260 nm. CD measurements were made using a JASCO J-715 spectropolarimeter fitted with a Peltier temperature controller.

The CD data shown in Fig. 3 provides spectroscopic evidence for the formation of helical structures by the SAF peptides. Fig. 3(A) shows circular dichroism (CD) spectra at 10 μM for: SAF-p1 (----), SAF-p2 (- -), the average of these spectra (- - -), and the experimental SAF peptide mixture (o). Fig 3(B) shows CD spectra at 100 μM - the key is the same as for Fig 3(A), but with the additional spectrum (•) being for the SAF peptide mixture after "maturation" for 1 h.

Consistent with our design, neither SAF-p1 nor SAF-p2 was highly structured in aqueous solution at pH 7 and 5 °C (Fig. 3). However, when mixed in equal proportions the circular dichroism (CD) spectrum changed and, moreover, was markedly different from the theoretical spectrum generated by averaging the spectra for the isolated peptides. In particular, the spectrum for the mixture had intense minima at 208 and 222 nm consistent with the formation of α-helical structure, but these features were not as pronounced in the spectra of the individual peptides. This was clear evidence that the two peptides interacted to form an α-helical structure as designed. Furthermore, and as expected for a multimerization event, the magnitude of these spectral changes depended on peptide concentration; a SAF mixture with 10 μM of each peptide, did show a weak signal indicative of some α-helical structure, however, a 100 μM mixture gave a much stronger signal (Figs. 3A&B).

The shape and intensity of spectra from 100 μM mixtures of the SAF peptides also changed with time (Fig. 3B). Spectra recorded immediately after mixing a "fresh" sample displayed some α-helical structure. After incubation of the mixture for 1 hour at 5 °C ("maturation"), however, the signal at 222 nm was more intense, and indicated approximately 75 % α-helix, consistent with substantial coiled-coil formation.

Maturation of 100 μM SAF peptide mixtures was also accompanied by slight clouding of the samples. Scattering effects from such samples can lead to attenuation and distortion of CD spectra (D. Mao and B. A. Wallace, (1984) *Biochemistry* 23, 2667). However, we could

disregard this possibility because altering the distance between the sample and the detector in the CD instrument did not affect the shape or the intensity of the spectrum. Furthermore, we established that the majority of the CD signal from the mixtures derived from the suspended material: a supernatant without the suspended material, which was recovered by centrifugation of a matured 100 μM SAF mixture, gave only a weak CD signal similar to the 10 μM mixture.

Thus, the CD data were wholly consistent with the desired α-helical SAF design and, moreover, indicated the formation of large assemblies.

As a control, SAF-p3 (the permutation of SAF-p1 (identical to SAF-p1c)) was designed to form a blunt-ended heterodimer with SAF-p1 that should not assemble further into fibres. 100 μM mixtures of SAF-p2 (identical to SAF-p2D) and SAF-p3 were analysed by sedimentation equilibrium in the analytical ultracentrifuge. The resulting data were best fitted assuming a single ideal species in solution, and the molecular weight was allowed to vary during the fit. An M_r of 6422 (with 95% confidence limits of 5924 and 6911) was obtained, which is very close to the expected heterodimer value of 6303 calculated from mass spectrometry of the individual peptides. CD spectra for 100 μM fibre-producing mixtures (SAF-p1 with SAF-p2), and for blunt dimer-producing mixtures (SAF-p2 with SAF-p3), were recorded. For the blunt dimer-producing mixtures, the shape and intensity of the CD spectrum were fully consistent with coiled-coil formation as designed. In contrast to the fibre-producing mixtures, the blunt dimer-producing mixtures showed no signs of maturation; that is, negligible spectral changes and no clouding of solutions occurred upon incubation. Interestingly, the intensity of the minimum near 222 nm, which is an accepted indicator of α-helical structure and degree of α-helical folding, was similar for both mixtures. This strongly supports the formation of α-helical structure as designed in the fibre-producing mixtures despite the spectral shifts observed upon maturation.

4) Linear Dichroism Experiments

Linear dichroism (LD) spectroscopy was also used to test if elongated structures were being formed as designed. Long polymers such as DNA molecules can be oriented by shear flow. This effect can be monitored by LD spectroscopy provided that chromophores also become

aligned by the flow (M. Bloemendal (1994) *Chem. Soc. Rev.* 23, 265; A. Rodger and B. Norden (1997) Oxford Chemistry Masters (Oxford University Press, Oxford), vol. 1).

Peptide samples were prepared for LD as for CD. LD data were collected on samples spinning in a couette flow cell by integrating the signal for 2 s every nm in the range 210-320 nm, using a JASCO J-715 spectropolarimeter. After baseline correction, absorbance was converted to molar extinction coefficient (1 mol-res^{-1} cm^{-1}) by normalizing for the concentration of peptide bonds. A linear correction for a sloping baseline was made to the data from the 100 μM SAF peptide mixture.

The results are depicted in Fig. 3D, which shows linear dichroism (LD) spectra for: 20 μM tropomyosin (- - -), the SAF peptide mixture at 10 μM (----), and the SAF peptide mixture at 100 μM in the absence (•) and presence (o) of 0.5 M KF.

For instance, we found that tropomyosin, which forms a dimeric coiled coil approximately 42 nm in length, could be aligned to give a LD signal (Fig. 3D). In contrast and consistent with our design and the CD data, experiments with a 10 μM SAF mixture, (Fig. 3D), and for the individual peptides at 100 μM (data not shown), LD signals were not detected. However, a matured 100 μM SAF peptide mixture gave a strong absorbance from the peptide backbone (210-240 nm) and some signal in the aromatic region (260-290 nm) during flow orientation (Fig. 3D). As only long structures are aligned by this technique, the data demonstrated that long fibres at least 500 nm in length were present in solutions of the matured 100 μM SAF peptide mixtures.

5) Electron Microscopy

To confirm fibre assembly, we used electron microscopy to visualize structures in the peptide preparations directly. For TEM experiments, peptide samples were incubated for 1 h at 5 °C in filtered 10 mM MOPS, pH 7. A drop of peptide solution was applied to a carbon-coated copper specimen grid (Agar Scientific Ltd, Stansted, UK), and dried with filter paper before negative staining with 0.5% aqueous uranyl acetate and then dried at 5 °C. A "fresh" SAF peptide mixture was prepared by mixing preincubated solutions of the individual peptides at 200 μM directly on the specimen grid, before drying and negative staining as described. Grids were examined in a Hitachi 7100 TEM at 100 kV and digital images were acquired

with a (800 x 1200 pixel) charge-coupled device camera (Digital Pixel Co. Ltd., Brighton, UK) and analyzed (Kinetic Imaging Ltd., Liverpool, UK).

For scanning electron microscopy (SEM) experiments, negatively-stained specimen grids were sputter-coated with gold and examined in a Leo Stereoscan 420 SEM at 20 kV and with a probe current of 10 pA.

No structures were visible up to 100 000 times magnification by transmission electron microscopy (TEM) for either the 10 μM SAF mixture, or for the individual peptides at 100 μM concentration (data not shown). However, TEM of a 100 μM SAF mixture at 50 000 times magnification revealed time-dependent formation of long fibrous structures, consistent with the CD and LD data. Fresh mixtures showed large numbers of extended fibres of various widths. The majority of these had a diameter of about 20 nm (Figs. 4A (a fresh mixture at 100 μM) & Fig 4D); finer fibres were present, but their widths could not be measured reliably. Images recorded for the matured mixtures showed fewer fibres, but these were more distinct and thicker than those observed in the fresh mixture (Fig. 4B&D).

Scanning electron microscopy (SEM) of a matured mixture showed no evidence for fibre branching. Rather, the fibres were simply intertwined as if layered on top of each other (Fig. 4C). It was not possible to follow the full length of fibres due to intertwining, but they were at least several hundred microns in length. Although the density of fibres varied across the surface of the EM grid, for the matured samples at least, their diameters were quite uniform with a mean width of 43.3 (SD = 9.3) nm (Fig. 4D). As the original design was for a longitudinally extended, but otherwise two-stranded coiled coil the average diameter that we might have expected was about 2 nm. Therefore, the EM data suggested that the designed two-stranded coiled-coil fibres associate laterally into higher order assemblies.

6) X-ray Fibre Diffraction

Mixtures of SAF peptides at 500 μM in 10 mM MOPS, pH 7, were incubated on ice for at least 1h, before centrifugation at 6500g for 5 min. Droplets of fibre-containing solutions, taken from the bottom of the centrifuged tubes, were suspended between the ends of two wax-filled capillaries and allowed to dry slowly overnight at 4°C, yielding clumps of partially aligned fibres. X-ray fibre diffraction images were collected using a Rigaku CuKα rotating

anode source (wavelength 1.5418 Å) and a R-AXIS IV detector. Samples were maintained at 5°C during data collection with cool air from a cryostream (Oxford Cryo-systems). The X-ray fibre diffraction pattern collected from SAF peptide fibres showed the following features (Fig. 6): (1) a short meridional (that is, parallel to the long fibre axis) reflection at 5.11± 0.03 Å; (2) the harmonic of this 5.11 reflection at 10.19 ± 0.05 Å; and (3) a stronger, more diffuse reflection centered at 8.8 ± 0.15 Å on the equator. These features are consistent with α-helical coiled-coils aligned with the fibre axis. The 5.1 Å meridional reflection corresponds to the pitch of the helices within the coiled-coils. The other expected reflection on the meridian-that is, that at 1.5 Å and corresponding to the rise per residue-lies out of the resolution of the current data sets, whereas the equatorial reflection reveals the mean distance between α-helical axes. This value at 8.8 Å is less than the observed value for keratin but falls within reported ranges for dimeric coiled-coil peptides.

7) Effect of Potassium Fluoride on Protein Fibre Assembly

Molecular modeling of the SAF sequences into an extended two-stranded coiled coil also highlighted potential complementary charge interactions on the surface of the protofibrils, Figs 1&2. In accordance with this, experimentally it was found that moderate concentrations of salt inhibited protofibril and thick fibre assembly. First, CD spectra recorded for both the individual peptides and a 100 μM mixture of SAF peptide samples with 0.5 M potassium fluoride showed reduced helical CD signals and there was no evidence of "maturing" in the mixed samples (Fig. 3C). Second, the LD signal described previously for the matured 100 μM SAF peptide mixture was also lost when the experiment was repeated in the presence of salt (Fig. 3D). Finally, TEM images of a 100 μM SAF mixture also demonstrated that fibres were not formed in the presence of 0.5 M KF (Fig. 4E). Fig. 4E shows the results of TEM of a matured SAF peptide mixture at 100 μM incubated in the presence of 0.5 M KF.

The inventors did not knowingly design any features into the SAF peptides to foster further association of the two-stranded coiled coils. The observation of thick fibres in SAF peptide preparations, therefore, raised the question: what interactions guided and stabilized these higher-order assemblies? The inventors therefore propose that features inherent in repeating structures of the type that they designed will naturally promote such fibre assembly (fibrillogenesis).

Consider a protofibril as depicted in Fig. 1B and 2A. Any sequence feature presented on its surface by either, or both of the constituent peptides will be repeated at regular intervals along the protofibril. The repeat length will be equal to the length of the peptides (for SAF-p1 and SAF-p2 this was 28 residues, or about 4.2 nm). Furthermore, the motif will spiral around the protofibril tracking the superhelix of the coiled coil, which has a pitch of about 15 nm for a contiguous, heptad-based, dimeric structure. In this scenario, protofibril-protofibril interactions may be promoted if another sequence motif complementary to the first is present in the potential partner. This is because the pitches of the complementary motifs on each protofibril will match precisely. Thus, once initiated, lateral association of protofibrils — that is, fibrillogenesis — will be cemented by many regularly spaced interactions as in a crystal. As a result, the complementary interactions need only be weak as the stability of the protofibril-protofibril interaction rests on an avidity effect rather than a small number of strong interactions. Provided that the components of the assembly can make more than one type of complementary surface very extensive molecular assemblies may result.

The inventors used electrostatic interactions both to direct heterodimer formation, and to promote elongation of the protofibrils (Figs. 1 and 2). These features would also create periodic and alternating patches of charge in the protofibrils provided they are regular as envisaged (Fig. 1B and 2B). These charged patches could guide and stabilize the higher order assemblies. Indeed, similar features have been noted in several natural fibrous proteins and have been implicated in the assembly of multi-protein filaments (J. J. Meng *et al* (1994) *Biol. Chem.* **269**, 18679; A. D. McLachlan and M. Stewart (1976) *Mol. Biol.* **103**, 271), and small synthetic peptide systems (S. G. Zhang *et al* (1993) *Proc. Natl. Sc. U.S.A* **90**, 3334). The experiments with salt (KF) described above suggest that salt-bridges (electrostatic interaction) may be at least in part the cause of fibrillogenesis.

8) Coiled-coils design

a. For two superimposed heptads there are three possible sequence offsets of 1, 2 and 3 residue(s), which are equivalent to 6, 5 and 4-residue offsets, respectively. For a regular 3.6-residue-per-turn α-helix, these set up two hydrophobic faces with angular offsets of 100°, 160° (360-200) and 60° (360-300), respectively, around the outside of the helix. This is best seen on a helical wheel. Accounting for helical supercoiling - i.e assuming 3.5 residues per turn and using the accepted helical-wheel representation for the

coiled-coil these angular offsets are altered to 103°, 154° and 51°, respectively. However, both sets of angles are over-simplifications when considering helix-helix interactions in actual coiled-coil systems because side-chain size, geometry and packing also affect the helix interfaces (Harbury, P. B. *et al* (1993) Science **262**, 1401-1407; Harbury, P. B. *et al* (1994) Nature **371**, 80-83; Malashkevich, V. N. *et al* (1996) Science **274**, 761-765). Nonetheless, we found that many natural coiled-coil assemblies, at least, were consistent with the approximate angular offsets: Trimers could be considered as having overlapping heptads separated by 3 residues (angular offset = 51/60°). Whereas, tetrameric and pentameric coiled-coils were often variations on a theme with two hepad repeats offset by 1 residue (100/103°).

b. **Two heptad repeats offset by two residues: α-cylinder constructions**

Sequence offsets of 2 residues are potentially more interesting than the 1- and 3-residue offsets. This is because of the possibility of placing hydrophobic (H) residues at a, c, d, and f, with c and f effectively making up the a' and d' positions of the second, offset heptad. This is represented below, where P signifies polar (non-core) residues.

a b c d e f g a b c d e f g	repeat 1
H P P H P P P H P P H P P P	binary pattern 1
P P H P P H P P P H P P H P	binary pattern 2
f'g'a'b'c'd'e'f'g'a'b'c'd'e'	repeat 2
a b c d e f g a b c d e f g	assigned register
H P H H P H P H P H H P H P	overall binary pattern

Such sequence patterns would results in two hydrophobic seams with a wide angular separation (154/160°), which would place them roughly on opposite sides of the helix. Furthermore, it offers two possibilities for parallel helix-helix packing arrangements: *syn*, where two like faces - i.e a / d with a / d , or c/f with c/f - from neighbouring helices combine to produce an openα-sheet, Fig. 6a; anti, where a/d faces pair with c/f. In the anti-arrangement the structure can close to form α α-cylinder. For antiparallel pairs of helices syn-typic association should lead to cylinders, whereas sheets should be formed

from anti-typic antiparallel interfaces.

c. **A natural α-cylinder**

TolC has two α-barrel-like domains (Koronakis, V. *et al* (2000) Nature **405**, 914-919). Both have 12 helices contributed by 3 monomers. In the lower barrel each helix pairs with another from the same protomer to form separate supercoiled, antiparallel coiled-coils; SOCKET analysis revealed extensive antiparallel knobs into holes (KIH) interactions within these pairs, but not between them. In contrast, the helices of the upper barrel appear to pack more uniformly, albeit with a slant, to describe an α-cylinder. The SOCKET output for this part of the structure revealed many fewer KIH interactions than found in the lower barrel. Furthermore, KIH interactions were not contiguous around the cylinder and, in particular, they were more extensive between helices in the same monomer, but less regular between the helices abutting the monomers. In our view, the TolC barrel represents a variation of the cylinders formed by protein structures of the invention.

Nevertheless, the inventors were able to assign heptad registers for the helices of the upper barrel unambiguously. This revealed knobs at relative *a*, *c*, *d*, and *f* positions and syn-typic association of two seams adjacent helices; i.e fully consistent with the theory outlined above.

We believe that it will be possible to constuct α-sheets and α-cylinders using helices in parallel. The use of parallel helices does have one interesting consequence for the construction of α-cylinders, however: as the pairing in these structures will be anti-typic, *a* residues on one helix partner *c* residues of a neighbouring helix at the same level in the structure. Similarly, *d* and *f* residues pair at the intervening levels. The result will be that successive helices will be translated up the helix and cylinder axes by two residues, which is equivalent to ≈3Å. Thus, attempts to construct α-cylinders from parallel helices will give spirals of helices which may or may not close. This is, however, potentially extremely interesting as it opens up possibilities for making peptide-based nanotubes as described above.

A second consideration for α-cylinder construction is the consequences of helix and coiled-coil supercoiling. The upper barrel of TolC has 12 helices. Based on a structure of parallel helices with canonical supercoiling, i.e an angular separation of 154° between the two seams in each helix, we calculated that the cylinder should close at 14 helices. However, variations in helix number are expected. One reason for this is that helices cannot supercoil in two direction simultaneously, and some distortion is required to maintain packing at both interfaces. We found structural precedents for this in the Protein Data Book PDB where tight knobs-into-holes packing was maintained (Walshaw & Woolfson, unpublished); indeed, the central helices of the 3-helix α-sheets are straight, Fig. 7b. (n.b . The slanting of the helices in the upper barrel of TolC may offer a compromise between straight and supercoiled helices). Assuming the packing of completely straight helices, the angular offset becomes 160° and 18 helices would close a cylinder. However, given that, as in 3-, 4- and 5-stranded coiled coils, side chains mediate the helix-helix contact angles other oligonmerisation states might be possible (Harbury, P. B *et al* (1993) Science **262**, 1401-1407; Harbury, P. B. *et al* (1994) **371**, 80-83; Malashkevich, V. N. *et al* (1996) Science **274**, 761-765): we calculate that small adjustments in the angular offset between 144° to 162° varies the helix number from 10 to 20.

9) Formation of Protein Structures

As mentioned above, the protein structures of the invention may have various applications such as in:

Nanotubes

a. This can be achieved for example by combining the aforementioned 7- and 11-residue repeats with offsets in the sequence. The effect would be eliminate the overall hydrophobic displacement. In other words, alternating heptad and hendecad repeats give an 18-residue repeat to match the α-helical repeat; in the α-helix, 18 residues span 5 helical turns exactly. It may therefore be possible to create a completely closed peptide nantotube (Fig. 5 shows part of a nanotube) In the parallel, straight helix case there would be 18 helices per turn of the "cylinder", and the rise per turn is 36 residues. Thus, a 36-residue peptide with a 7-11-7-11 repeat offset by 2 residues should form a spiral of

helices the ends of which meet to close the tube. Such nanotubes maybe particularly useful in the production of nanoscale piping and plumbing. The interior of the tube may be derivatised to control the flow of different small (2-40Å) molecules.

b. **Derivitised and branched peptides and peptide templates**

The self-assembling peptides of the invention are relatively small and synthetically accessible. Thus, non-standard derivatisable side chains may be incorporated in them. For example, the monomer units can be made with a single cysteine residue at an exterior f position. These can be used to couple small molecules and other peptides using thiol-based chemistry. A wide variety of thiol-reactive probes are available. In particular, the peptides can be tagged with fluorophores. For instance, with one peptide labelled with Fluorescein and the other with Rhodamine fibres visualised by confocal microscopy appear green and red, respectively (Fig. 7). There is a possibility for FRET between the probes, which may pack closely in the fibres, and this may confuse interpretation. To avoid this the tagged peptides can be doped into fresh, assembling SAF mixtures. Having available fluorescently labelled peptides and fibres offers another route to tracking fibre/network assembly and orientation.

To generate branched self-assembling fibres "T-shaped" conjugated peptides can be made. These are covalent heterodimers made by mixing and coupling together variants of two SAF peptides: one with a terminal cysteine and the other having a central cysteine residue. The desired products can be purified from the mix of disulphide-linked peptide by PHLC. Doping the conjugated ("T") peptides into fresh SAF mixtures should propagate fibre assembly in three dimensions as both the "bar" and the "stem" of the "T" could become incorporated in, or initiate, fibres. The resulting networks can be visualised and characterised by EM.

Peptide synthetic diblock copolymer hybrids may be produced. Suitable methods for preparing water soluble diblock copolymers using atom transfer radical polymerisation are described in X. S. Wang *et al* Chemical Communications 1817 (1999) and X. S Wang *et al* Macromolecules 33, **257** (2000).

The protein fibres of the invention may be used to template and control this

polymerisation either to produce hybrid fibres or if the peptide template is subsequently disassembled and marked away, to provide routes to water soluble "fishnet" nanotubes. Other possibilities include: conjugating polymers onto preassembled peptide fibres; conjugating the polymers and peptides prior to fibre assembly; and effecting polymerisation on the pre-assembled fibres.

c. **Formation of Matrices**

The protein fibres of the invention may be arranged to form two and three dimensional grids and matrices respectively. One application for such matices is in the purification of biological fluids such as blood. An affinity matrix could be assembled (for example *in situ* in blood) to remove blood contaminants such as viruses. In the case of virus removal, a binder for the target contaminant (e.g a peptide or protein with natural or engineered affinities for a viral coat protein) can be fused to a peptide monomer units in the protein structure of the invention. The matrix can then be removed from blood along with any bound contaminants by light centrifugation. For example, it is estimated that a 100 nm length of fibre would have a mass of $\geq$ 12 MDa which would readily be removed. Such affinity matrices have a number of advantages over larger naturally occurring proteins. In the assembled matrices any binders are aligned to give high effective avidities for the targeted molecules.

d. **Other applications**

Other applications for protein structures in accordance with the invention include:

i. preparation of organised networks for seeding the crystalisation of biomolecules for X-ray crystallography;
ii. using ordered fibres to promote cell growth for tissue engineering;
iii. the construction of nanoscale molecular sieves
iv. the preparation of nanoscale molecular grids/scaffolds that could be used as supports for a variety of functional small or macromolecules.
v. functionalised grids and networks could be used in, for example, catalysis, affinity-sieving/purification of biological fluids and other research solutions, the recruitment of endogenous molecules and co-factors to promote tissue repair and tissue engineering in general.

vi. to create novel lab-on-chip technologies, peptide self-assembly could be combined with lithography as follows.

Lithography and related techniques can be used to pattern a variety of surfaces with channels, which can be made of a suitable size (e.g 20-100 nm wide and deep) to accommodate peptide fibres. These can then be used to direct the assembly of the fibres from solutions mixed directly on the surfaces. Furthermore, using well-established chemistry, the inventors envisage funtionalising the peptide fibres with a variety of small molecules and other proteins. This proposed combination of peptide design, self-assembly and lithography should allow the development of ordered arrays of functional polymers on specific surfaces.

vii. Assembled fibres could also be used as fine (therefore, high resolution) tips in AFM (atomic force microscopy) the current limit is about 10-25 nm using carbon nanotubes.

Claims

1. A protein structure comprising a plurality of first peptide monomer units arranged in a first strand and a plurality of second peptide monomer units arranged in a second strand in which a first peptide monomer unit in the first strand extends beyond a corresponding second peptide monomer unit in the second strand in the direction of the strands.

2. A protein structure according to claim 1 in which the strands together form a coiled coil structure.

3. A protein structure according to claim 1 or 2 in which at least one charged amino acid residue of a first peptide monomer unit is arranged to attract an oppositely-charged amino acid residue of a second peptide monomer unit.

4. A protein structure according to claim 3 in which the charged amino acid residue is in an end portion of the first peptide monomer unit which extends beyond the corresponding second peptide monomer unit in the second strand.

5. A polypeptide structure according to any preceding claim in which at least one strand consists solely of first or second peptide monomer units respectively.

6. A protein structure according to any preceding claim in which the peptide monomer units comprise a repeating structural unit.

7. A protein structure according to claim 6 in which the repeating structural unit comprises a heptad repeat motif (abcdefg).

8. A protein structure according to claim 6 in which the repeating structural unit comprises a hendecad repeat motif (abcdefghijk)

9. A protein structure according to claim 6 having isoleucine or asparagine at position a and leucine at position d.

10. A protein structure according to claim 6 having valine or leucine at positions a and d respectively.

11. A protein structure according to any one of claims 7 to 10 having oppositely-charged or otherwise complementary residues at positions g and e of respective monomer units.

12. A protein structure according to claim 11 in which the oppositely-charged residues are glutamic acid and lysine residues or asparagine and aspartic acid residues, or synthetic derivatives of these amino acid residues.

13. A protein structure according to any preceding claim in which the structure is stabilised by pairs of asparagine, arginine, lysine or other complementary residues provided by corresponding first and second peptide monomer units.

14. A protein structure according to any preceding claim which is arranged to form a tubular structure.

15. A protein structure according to claim 14 in which the peptide monomer units are offset by two or more amino acid positions in sequence whereby the peptide monomer units form a cylinder.

16. A protein structure according to any preceding claim in which the first and second peptide monomer units have the sequence:

a) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1) and

b) KIRALKAKNAHLKQEIAALEQEIAALEQ (SAF-p2) respectively; or

c) KIAALKQKIAALKQEIDALEYENDALEQ (SAF-p1A) and

d) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C) respectively; or

e) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1C) and

f) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C) respectively.

17. A peptide monomer unit for use in preparing a protein structure the peptide monomer unit having an amino acid sequence selected from:

a) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1);

b) KIRALKAKNAHLKQEIAALEQEIAALEQ (SAF-p2);

c) KIAALKQKIAALKQEIDALEYENDALEQ (SAF-p1A);

d) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C);

e) KIAALKQKIASLKQEIDALEYENDALEQ (SAF-p1C) ; and

d) KIRALKWKNAHLKQEIAALEQEIAALEQ (SAF-p2C).

18. A protein structure or peptide monomer unit according to any preceding claim in which at least one amino acid residue is derivatised.

19. A method of producing protein structures, the method comprising providing a mixture of first and second peptide monomer units which associate to form a protein structure according to any preceding claim.

20. A method according to claim 19 in which the protein structure is derivatised.

21. A method according to claim 19 or 29 in which the protein structure is stabilised by cross-linking.

22. Protein fibres produced by an association of protein structures according to any one of claims 1 to 3 or a method according to claim 19, 20 or 21.

23. A kit for making protein structures, the kit comprising first and second peptide monomer units which associate to form a protein structure according to any one of claims 1 to 13 or protein fibres according to claim 22.

24. A two dimensional matrix comprising a protein structure according to any one of claims 1 to 13 or protein fibres according to claim 22.

25. A three dimensional grid comprising a protein structure according to any one of claims 1 to 13 or protein fibres according to claim 21.

26. A matrix according to claim 25 which is arranged to assemble in solution.

27. A matrix according to claim 25 or 26 which is arranged to bind a target entity.

28. A matrix according to claim 27 which is arranged to bind viruses.

29. A method of forming a matrix according to any one of claims 25 to 28 in which a mixture of separate first and second monomer units is provided and are then caused to associate to form a protein structure in accordance with the invention, an accumluation of such protein structures assembling in turn to form a three dimensional matrix.

30. A method according to claim 29 in which the matrix is formed *in situ*.

31. A method for controlling the production of a synthetic polymers comprising assembling a protein structure in accordance to any one of claims 1 to 16 in association with the polymer.

32. A method according to claim 31 in which the protein structure is removed after synthesis of the polymer.

33. A tip for use in Atomic Force Microscopy comprising a protein structure according to any one of claims 1 to 16.

Peptides

This invention relates to protein fibre formation and in particular to methods of producing protein fibres to form a protein structure comprising a plurality of first polypeptide units arranged in a first polypeptide strand and a plurality of second polypeptide units arranged in a second polypeptide strand, the strands preferably forming a coiled coil structure, and in which a first polypeptide unit in the first strand extends beyond a corresponding second polypeptide unit in the second strand in the direction of the strands.

FIG. 1A

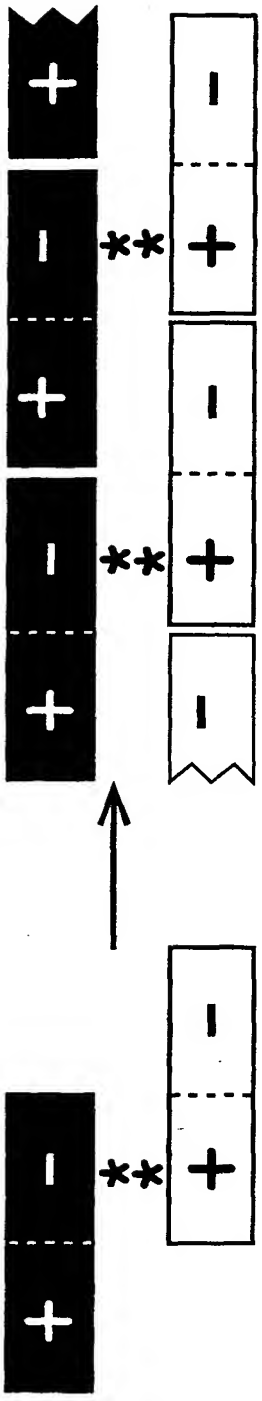


FIG. 1B

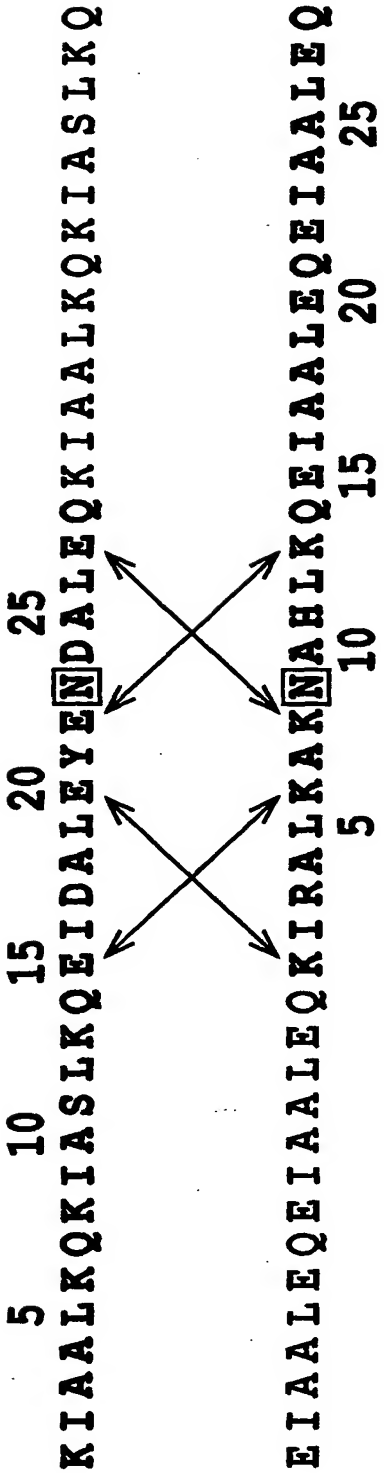
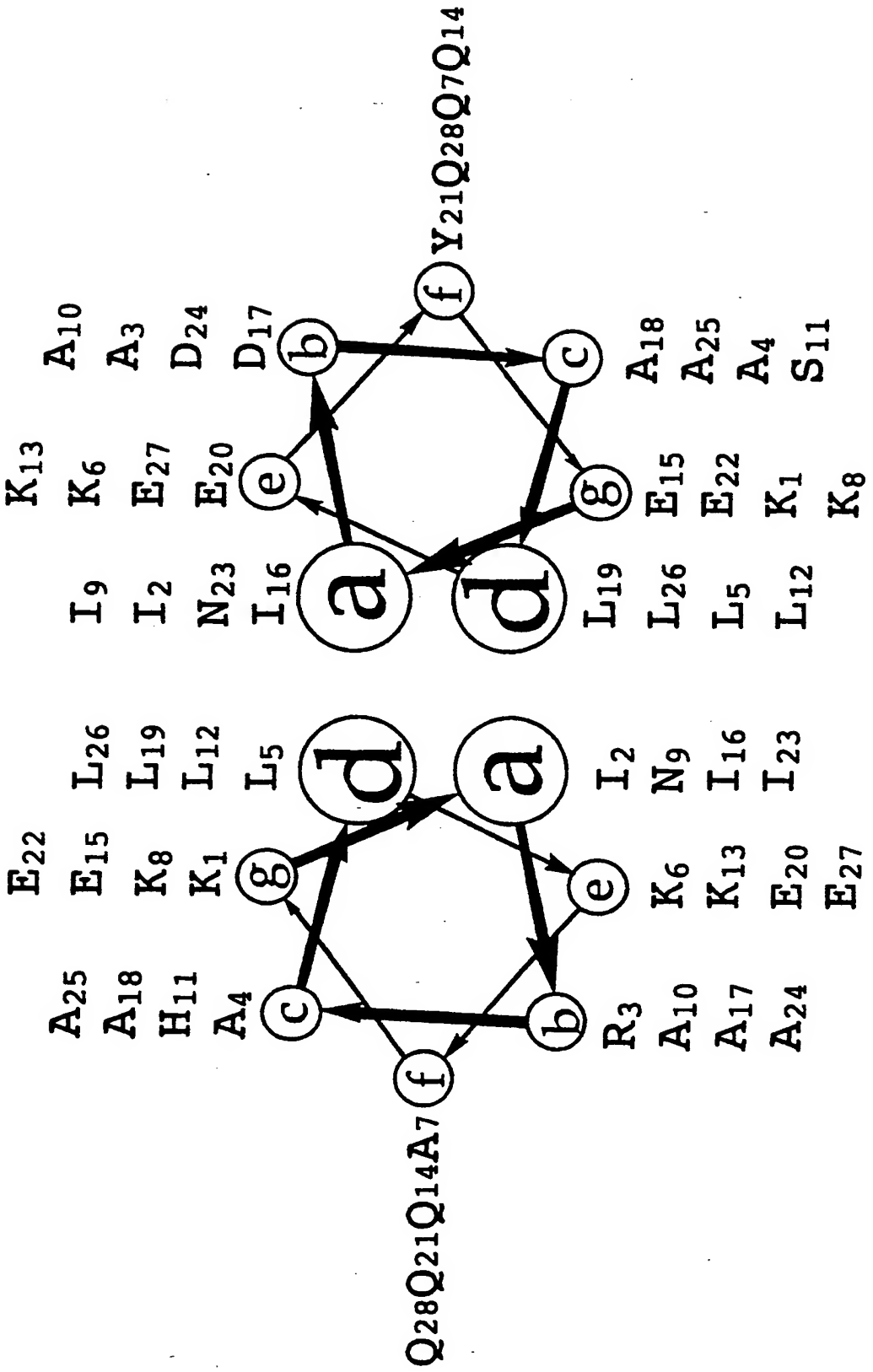

KIAALKQKIASLKQEIDALEYENDALEQKIAALKQKIASLKQ
5 10 15 20 25
EIAALEQEIAALEQKIRALKAKNAHLKQEIAALEQEIAALEQ
5 10 15 20 25

FIG. 1C


Y21Q28Q7Q14
A10 A3 D24 D17
K13 K6 E27 E20
I9 I2 N23 I16
L19 L26 L5 L12
A18 A25 A4 S11
E22 E15 K8 K1
Q28Q21Q14A7
A25 A18 H11 A4
E22 E15 K8 K1
L26 L19 L12 L5
I2 N9 I16 I23
R3 A10 A17 A24
K6 K13 E20 E27
a
d
b
c
e
f
g

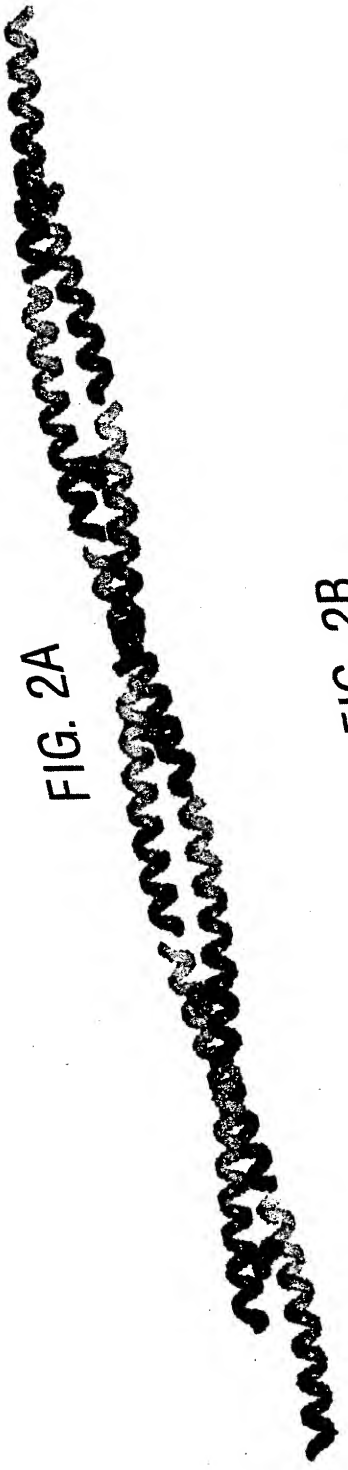


FIG. 2A

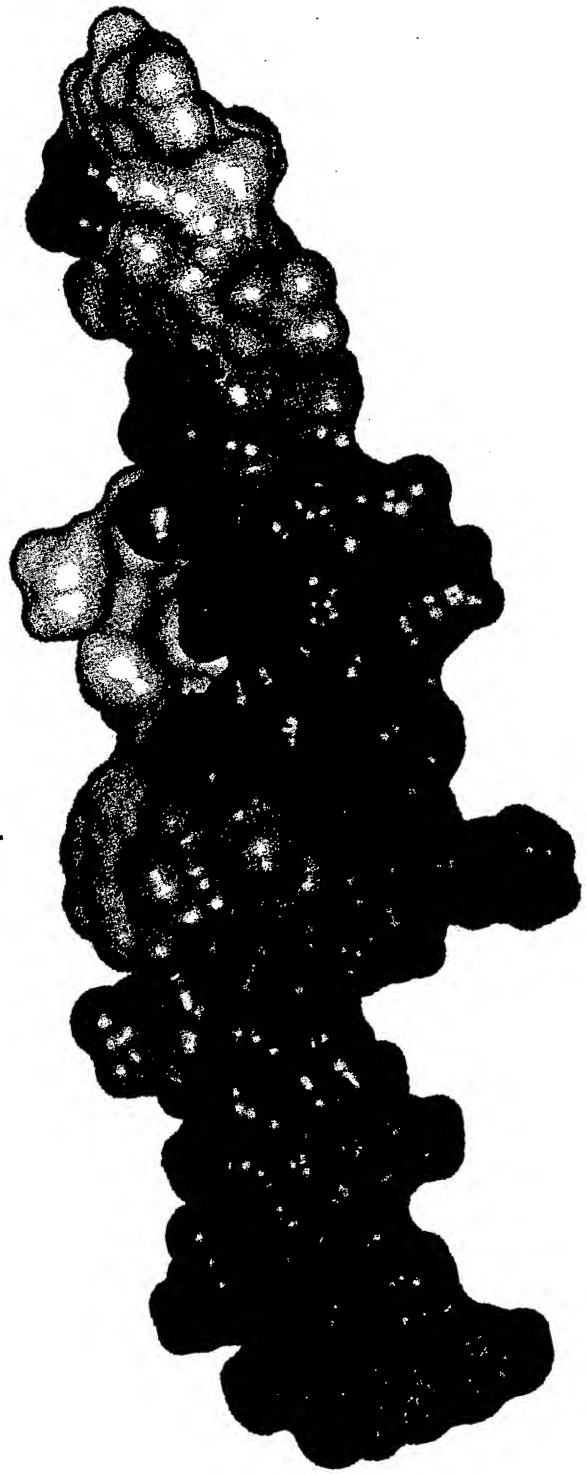


FIG. 2B

FIG. 3B

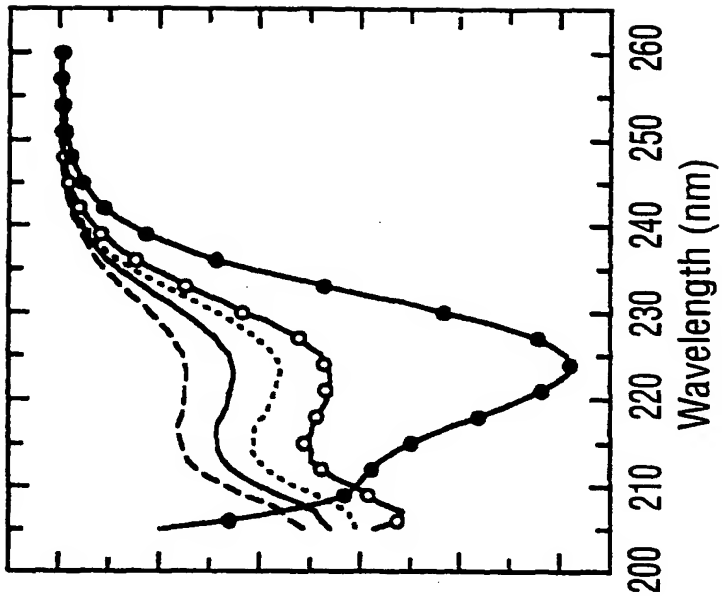

200
210
220
230
240
250
260
Wavelength (nm)

FIG. 3A

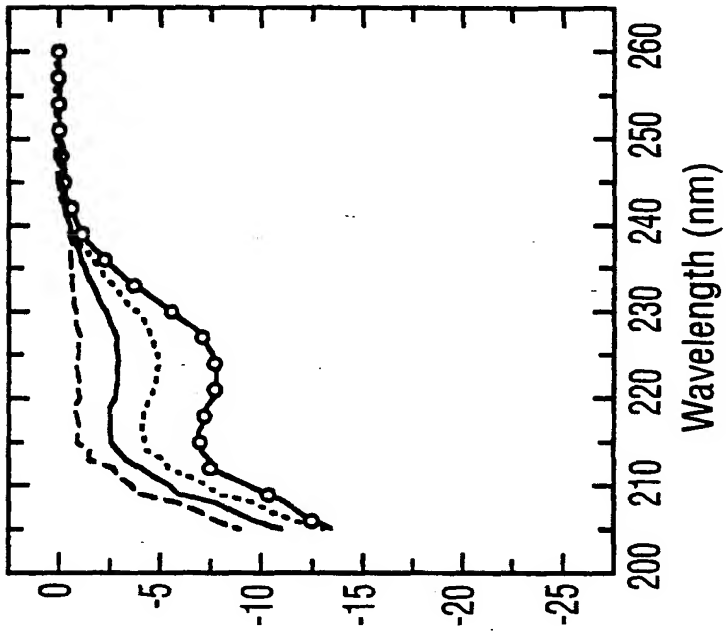

Molar Ellipticity
(deg cm² dmol-res⁻¹ x 10-³)
0
-5
-10
-15
-20
-25
200
210
220
230
240
250
260
Wavelength (nm)

FIG. 3D

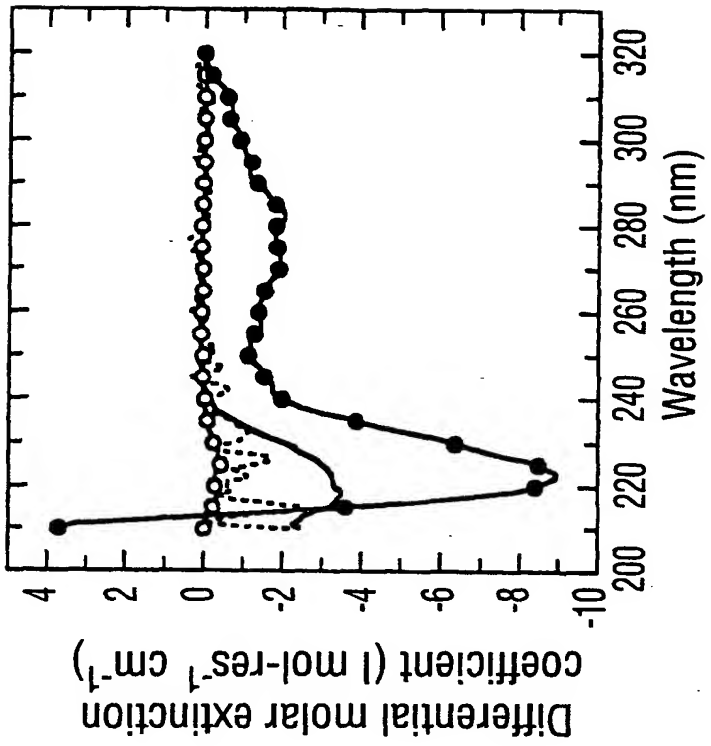

Differential molar extinction coefficient (l mol-res^{-1} cm^{-1})
4
2
0
-2
-4
-6
-8
-10
200
220
240
260
280
300
320
Wavelength (nm)

FIG. 3C

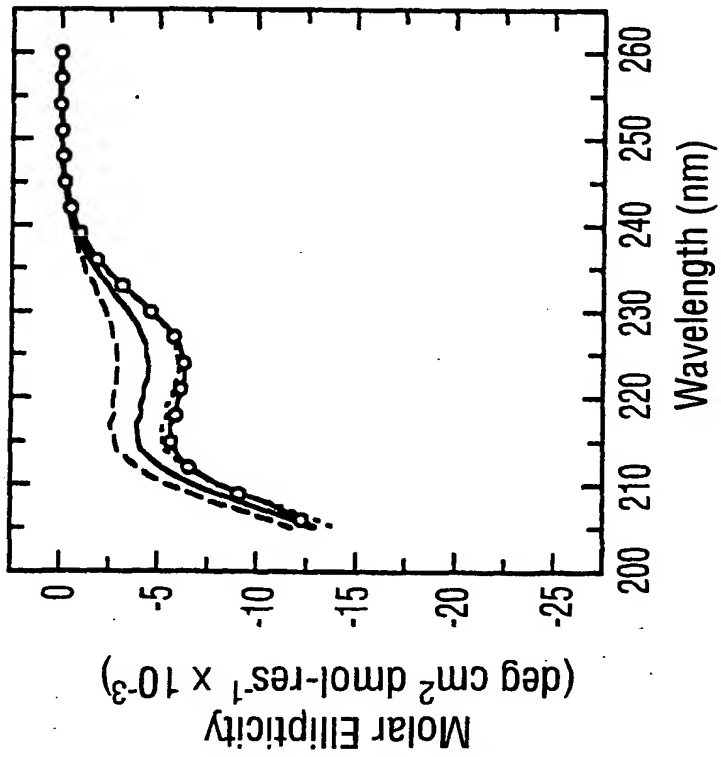

Molar Ellipticity (deg cm^2 dmol-res^{-1} x 10^{-3})
0
-5
-10
-15
-20
-25
200
210
220
230
240
250
260
Wavelength (nm)

FIG. 4B

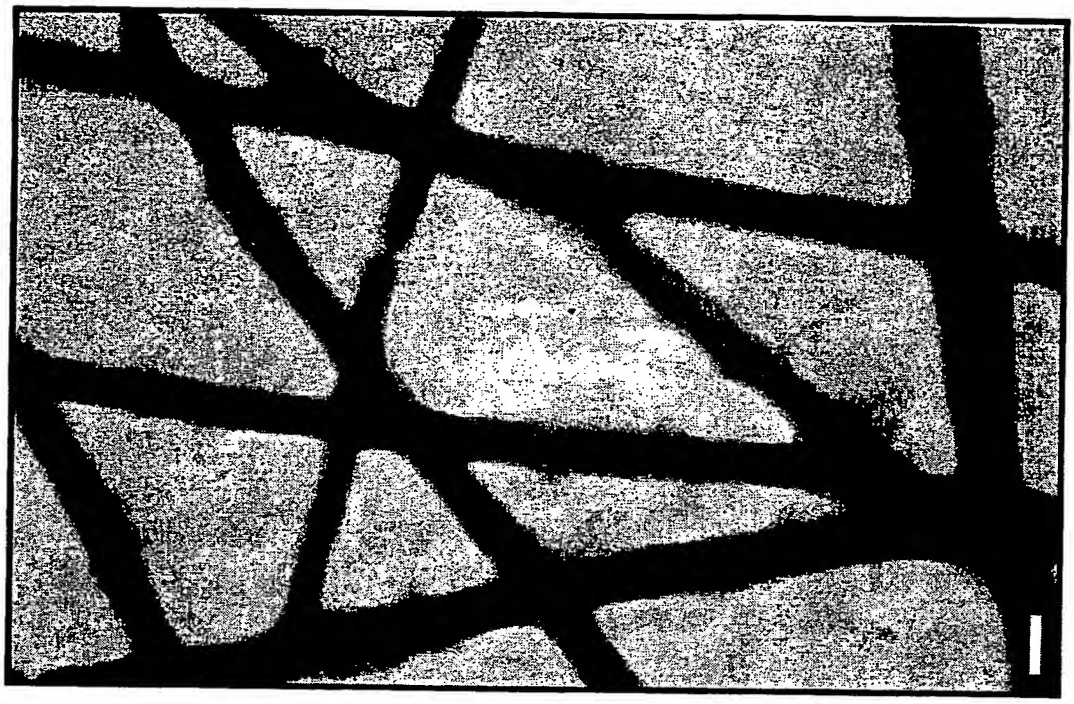


FIG. 4A

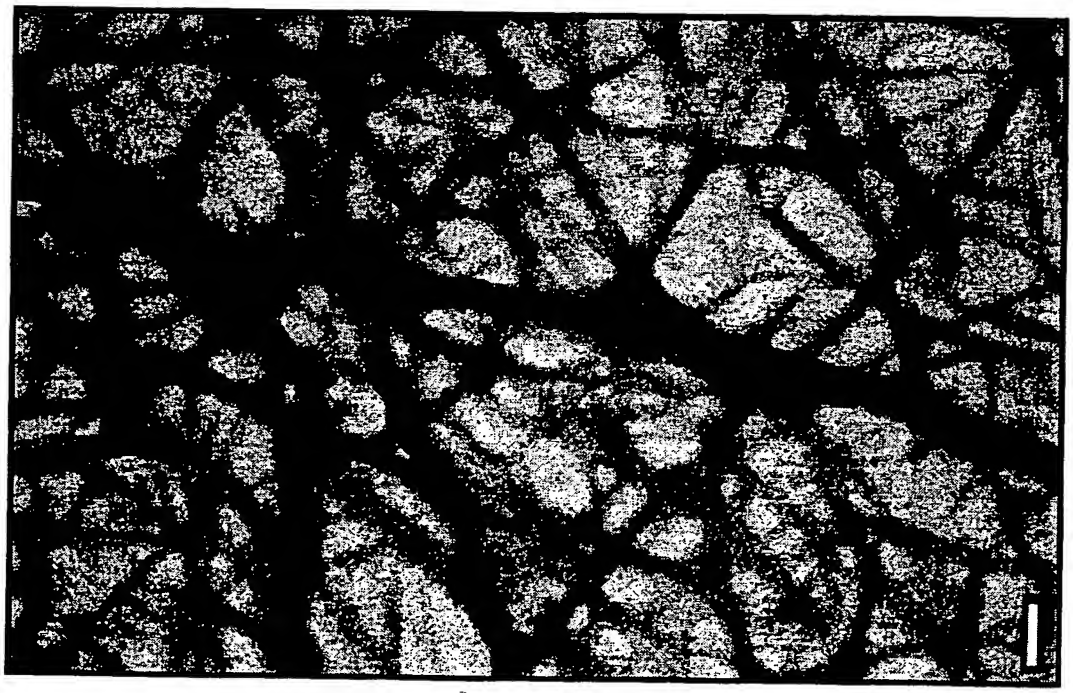


FIG. 4E

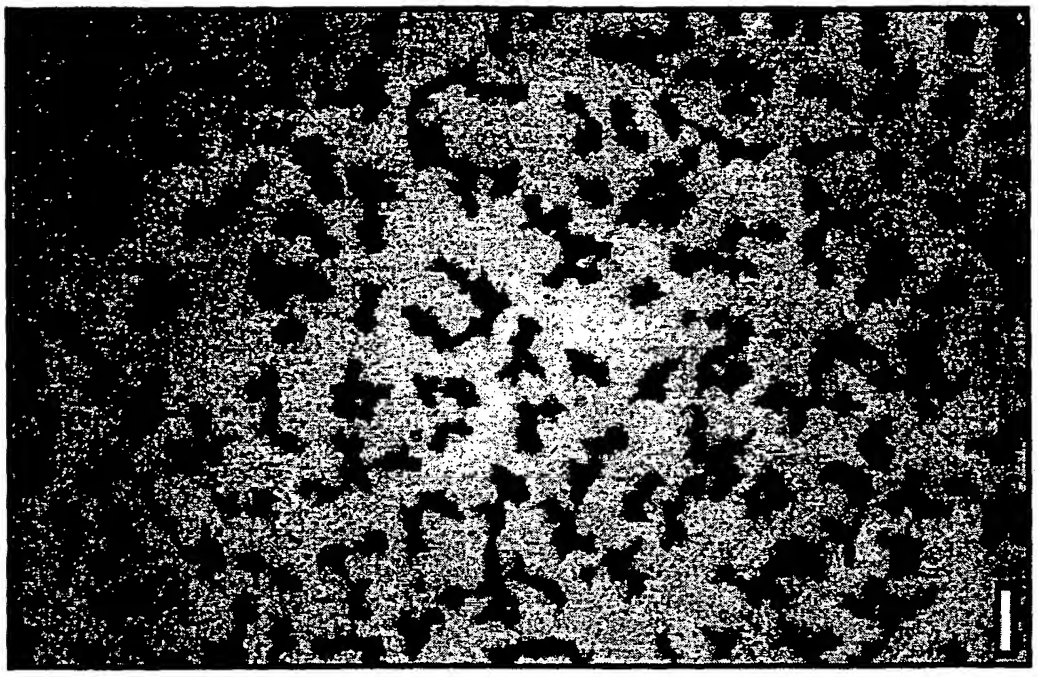
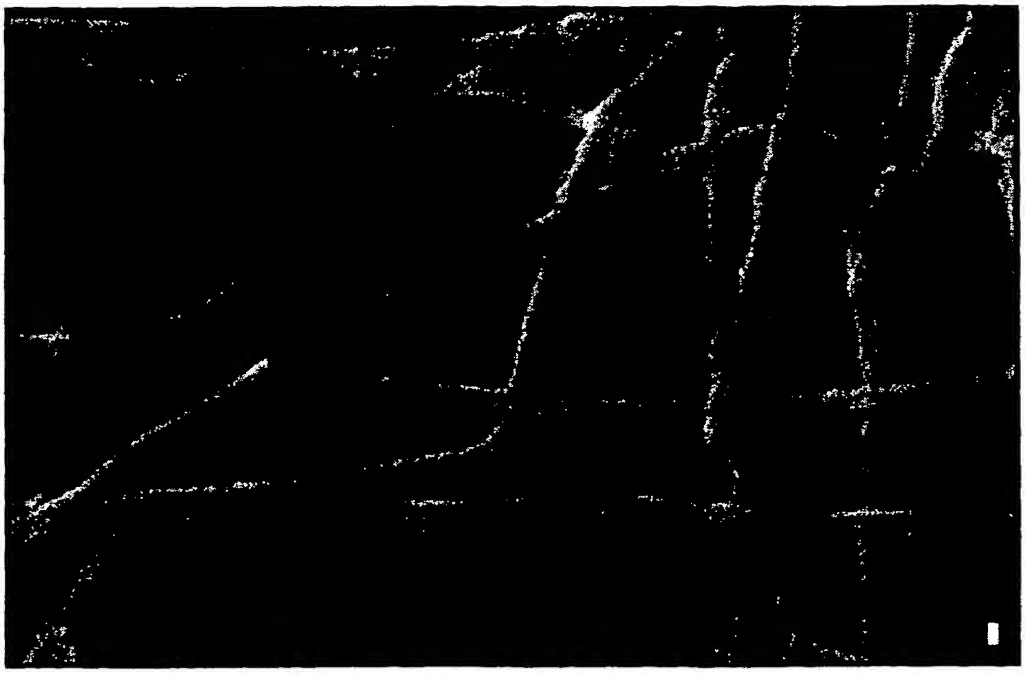


FIG. 4C



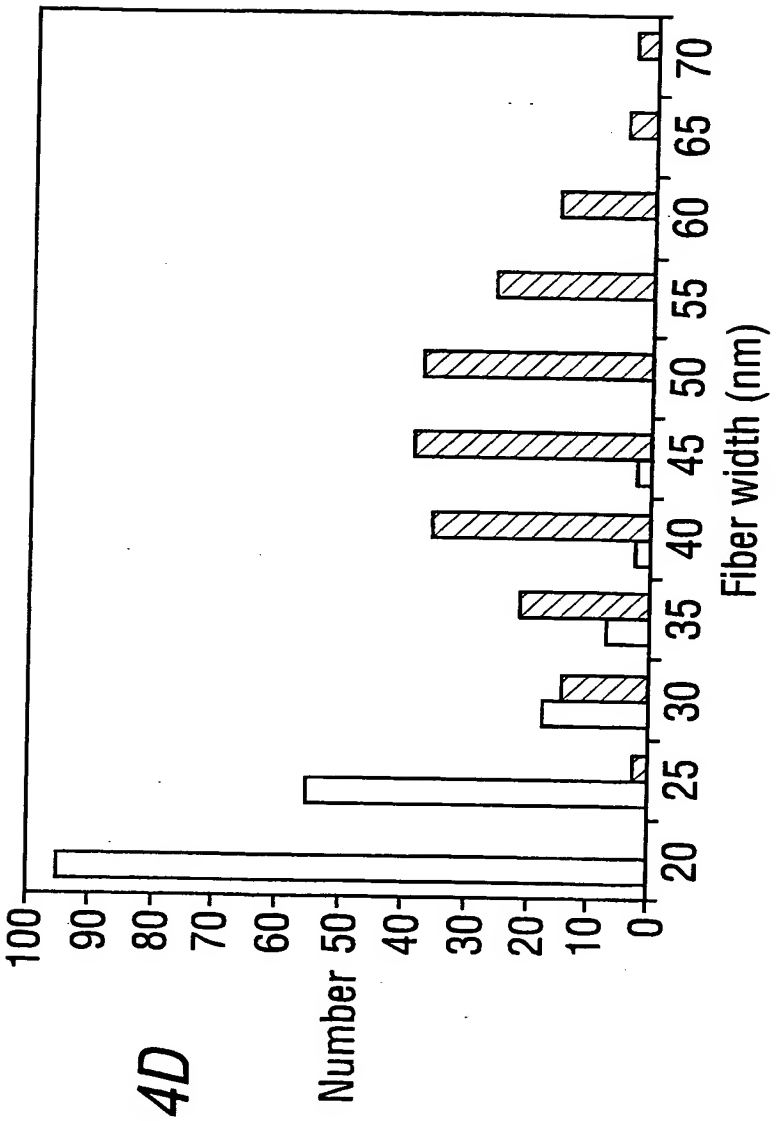
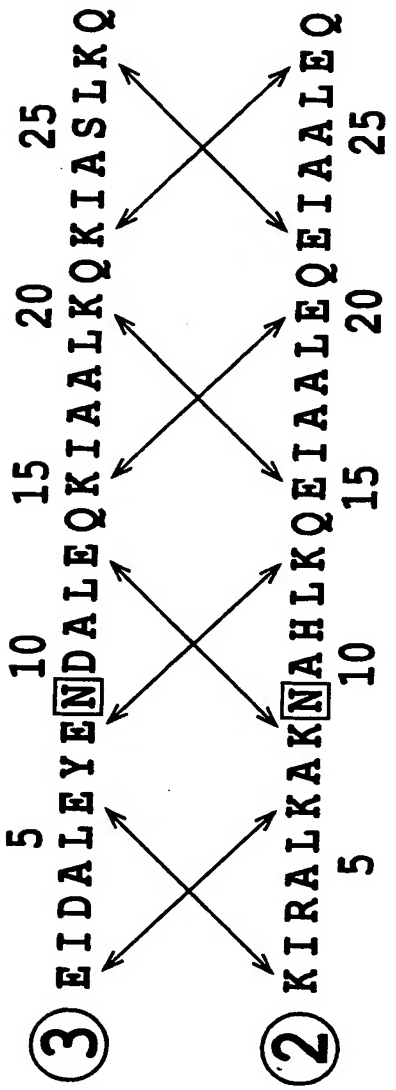

Number
Fiber width (nm)
0
10
20
30
40
50
60
70
80
90
100
20
25
30
35
40
45
50
55
60
65
70

FIG. 4D


③ EIDALEYENDALEQKIAALKQKIASLKQ
② KIRALKAKNAHLKQEIAALEQEIAALEQ
5
10
15
20
25

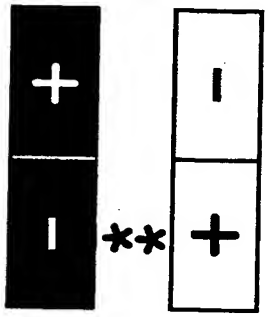

+
-
**
+
-

FIG. 8

FIG. 5

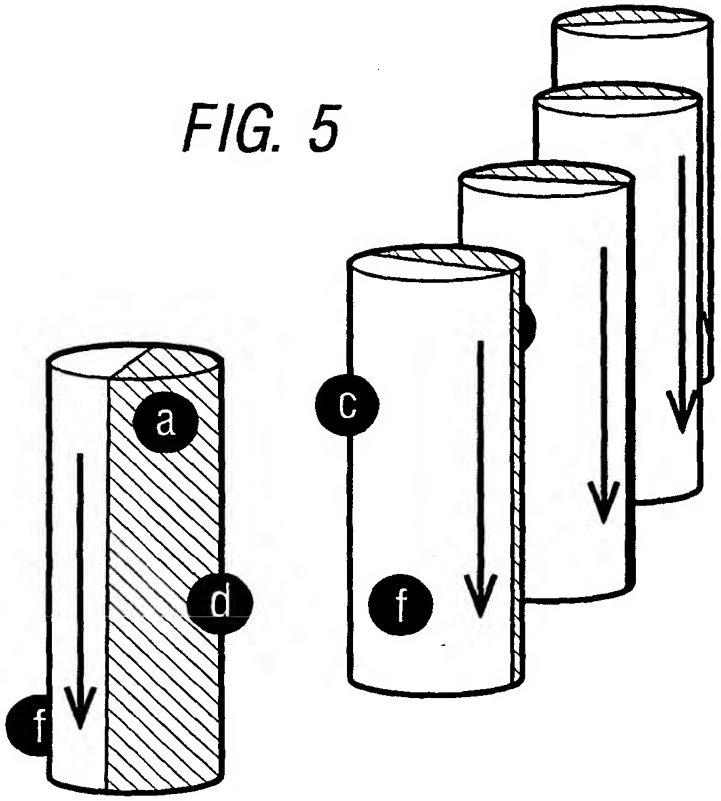
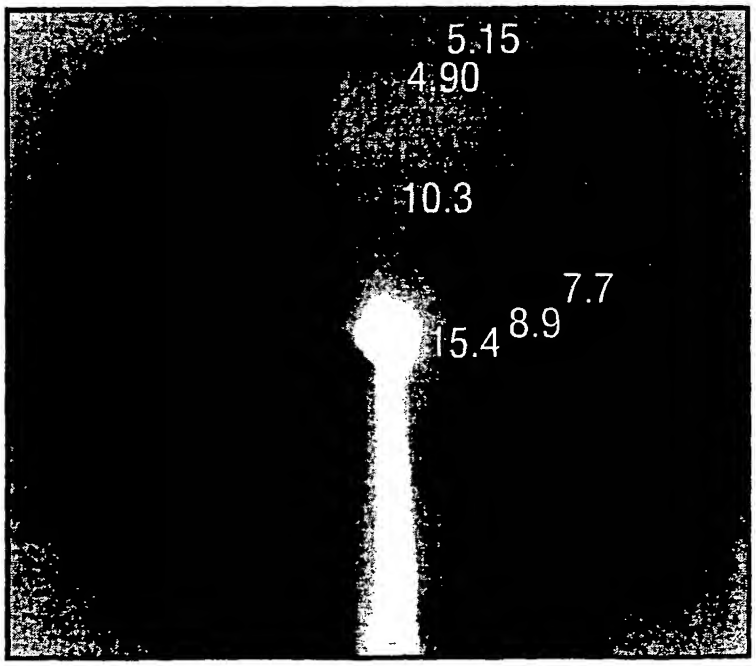


FIG. 6



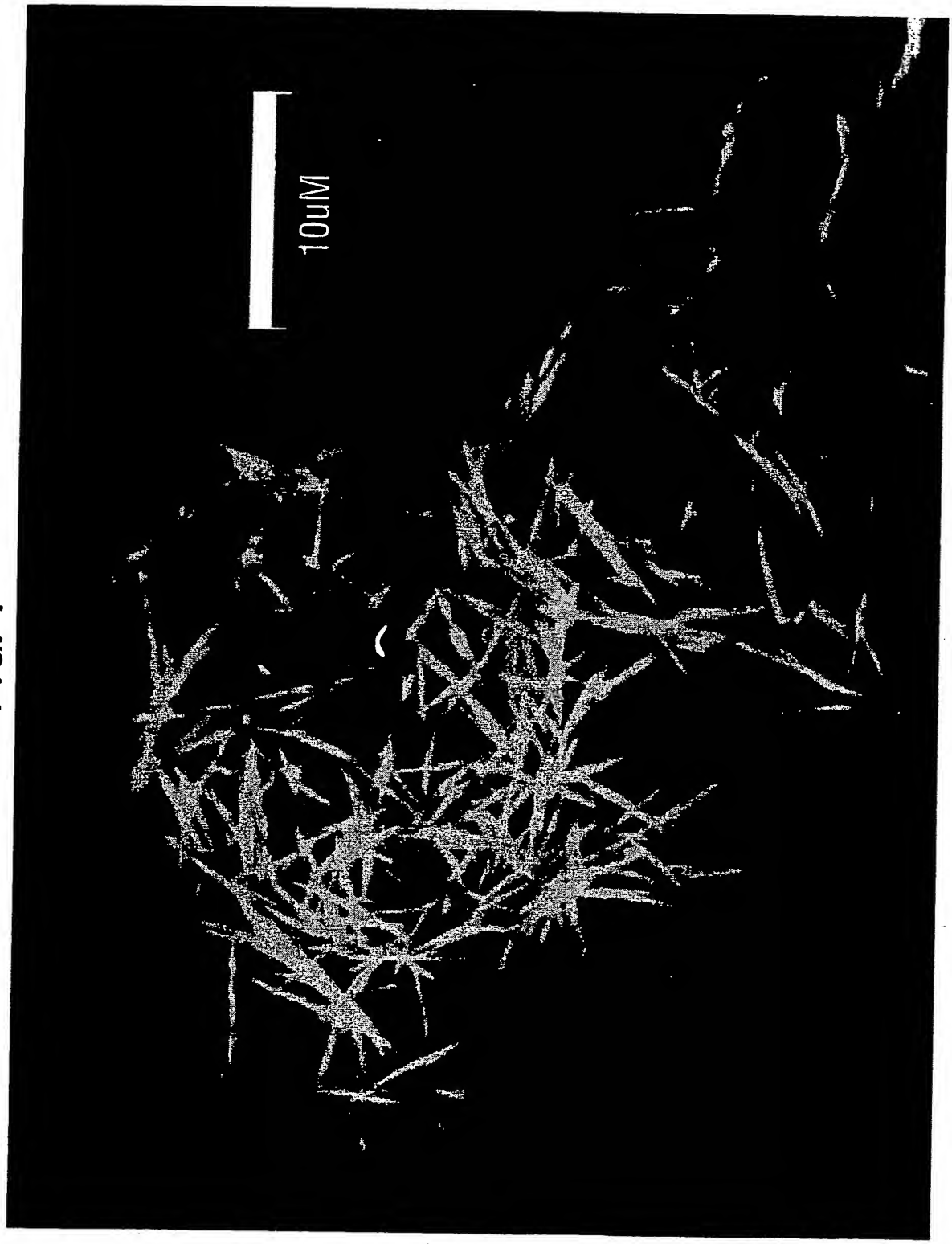


FIG. 7

INTERNATIONAL SEARCH REPORT

Intern. Application No
PCT/GB 00/03576

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/00 C07K1/113 G11B9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, WPI Data, PAJ, CHEM ABS Data, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 712 366 A (KAPLAN DAVID L ET AL) 27 January 1998 (1998-01-27) column 6, line 35-57; figures 6A,6B	1-15, 18-30
X	WO 96 11947 A (GOLDBERG EDWARD B) 25 April 1996 (1996-04-25) page 7-9; figures 3,4 page 25-26	1-6,14, 15, 18-20, 22-30

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 January 2001	17/01/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Cervigni, S

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Internal Application No
PCT/GB 00/03576

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	KOJIMA SHUICHI ET AL: "FIBRIL FORMATION BY AN AMPHIPATHIC ALPHA-HELIX-FORMING POLYPEPTIDE PRODUCED BY GENE ENGINEERING." PROCEEDINGS OF THE JAPAN ACADEMY SERIES B PHYSICAL AND BIOLOGICAL, vol. 73, no. 1, 1997, pages 7-11, XP000971780 1997 ISSN: 0386-2208 abstract	1-15,19, 22-30
X	W A PETKA ET AL: "REVERSIBLE HYDROGELS FROM SELF-ASSEMBLING ARTIFICIAL PROTEINS" SCIENCE,AAAS. LANCASTER, PA,US, vol. 281, 17 June 1998 (1998-06-17), pages 389-392, XP002149253 abstract; figure 2	1-13, 18-20, 23-30
A	KOHN W D ET AL: "De novo design of alpha-helical coiled coils and bundles: models for the development of protein-design principles" TRENDS IN BIOTECHNOLOGY,NL,ELSEVIER, AMSTERDAM, vol. 16, no. 9, September 1998 (1998-09), pages 379-389, XP004173181 ISSN: 0167-7799	

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern[redacted] Application No
PCT/GB 00/03576

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5712366 A	27-01-1998	NONE	
WO 9611947 A	25-04-1996	US 5877279 A	02-03-1999
		AU 689662 B	02-04-1998
		AU 3829695 A	06-05-1996
		BR 9509487 A	30-09-1997
		CA 2202474 A	25-04-1996
		CN 1168676 A	24-12-1997
		EP 0785946 A	30-07-1997
		HU 77683 A	28-07-1998
		JP 10508194 T	18-08-1998
		NZ 295046 A	25-03-1998
		US 5864013 A	26-01-1999